Approximations of Optimal Alarm Systems For Anomaly Detection

Rodney A Martin, Ph.D. *Member, IEEE*

***Abstract*— In many engineering systems, the ability to give an alarm prior to impending critical events is of great importance. These critical events may have varying degrees of severity, and in fact they may occur during normal system operation. In this article, we investigate approximations to theoretically optimal methods of designing such alarm systems for zero-mean linear dynamic systems driven by Gaussian noise. This simple modeling paradigm suffices due to the nature of the engineering and/or behavioral systems provided as examples to motivate the use of these methods.**

One example addresses thermal comfort applications for commercial buildings. Another example addresses integrated caution and warning health management systems for spacecraft propulsion. For both examples, an alarm may be given for any number of level-crossing events that occur over a specified time period. As such, an optimal alarm system can be designed to warn facility managers or ground-based telemetry data analysts of impending complaints or anomalous engine events, respectively. This will aid them in making critical decisions about building or spacecraft operations.

***Index Terms*— Optimal alarm theory, Level-crossing theory, Kalman prediction, Anomaly Detection**

I. Introduction & Background

THIS article introduces a novel approach of combining the practical appeal of Kalman prediction techniques with level-crossing theory and optimal alarm system design. A comprehensive demonstration of practical application for the design of optimal alarm systems has been covered in the literature [1], [2], [3]. However, the background theory for optimal alarm systems has seen modest coverage by other authors as well [4], [5], [6], [7]. The latter is by no means a comprehensive list, but illustrates a cross-section of the primary authors responsible for introducing optimal alarm systems in a classical and practical sense.

It was shown by Svensson [1], [2] that an optimal alarm system is fundamentally based upon a likelihood ratio criterion via the Neyman-Pearson lemma. This allows us to design an optimal alarm system that will elicit the fewest possible false alarms for a fixed detection probability. This becomes important when considering the numerous applications that might benefit from an intelligent tradeoff between false alarms and missed detections, by applying the theory and methodology introduced in this article.

There are several other decision rules that can be used from hypothesis testing/decision theory, in lieu of the Neyman-Pearson decision rule used as the basis of optimality here. Among these are Bayesian, MAP (maximum a posteriori), maximum likelihood, and the minimax criteria. The latter is derived from the Bayesian criterion, and seeks to minimize the maximum risk. Some recent interesting developments have even described adaptive on-line techniques using the Bayesian formulation [8]. However, there are still considerable computational issues, and a well-defined cost function is still required, even when the posterior probability is adaptively updated.

Here we present two contrasting examples representing distinct applications. The first example is based upon prediction and alarm of thermal sensation complaints in buildings, previously presented in [3]. As such, some of the technical details for this example will be presented in a more concise fashion. The second example is based upon fault detection and diagnostic work for spacecraft propulsion systems, as alluded to in [9]. However, both examples share the quantification of *any number* of level-crossing events that may occur over a specified time period. Both examples also assume quite liberally that the practical events of interest can sufficiently be characterized by this class of level-crossing events. That is, we assume that both thermal sensation complaints and spacecraft engine anomalies can accurately be represented by level-crossings, whose processes are characterized or transformed into zero-mean linear dynamic systems driven by Gaussian noise. More evidence to support this modeling paradigm will be presented subsequently.

Several examples of level-crossing events within this class will be studied here, varying from the simple case which involves two adjacent time slices, to the much more complicated case that involves a level crossing event that may span many time slices and exceed the level many times during this time-frame. The former more simple case is traditionally studied in the Swedish literature and invokes ARMA(X) prediction methods [1], [2], [4], [5], [6], [7]. A variant of the latter more complicated case has been investigated by Kerr [10] and uses a Kalman-filter-based approach.

There is an extensive history of invoking Kalman-filter-based approaches within the failure detection literature. A few of the most groundbreaking articles that discuss the use of Kalman filter methods for failure detection have been authored by Kerr [10], and Willsky and Jones [11]. Both of these articles have a long history of related methods descending from them, i.e., [12] which alludes to the use of the Neyman-Pearson lemma. However, these methods have not been without debate over the years, with one recent criticism of [11] addressing the claim of its optimality by Kerr [13].

The method presented by Willsky and Jones is charac-

Manuscript received xxx xx, 2007; revised xxx xx, xxxx. This work was supported by xxxx.

Rodney A. Martin is with the NASA Ames Research Center

terized by a formulation of the anomaly detection problem involving the GLR (generalized likelihood ratio) test. The method derived by Kerr shows how to derive a failure detection algorithm whose design is performed by computing false alarm and correct detection probabilities over a time interval. Both methods are related to, but not directly derived from optimal alarm system theory based upon level-crossings introduced by DeMaré [5] and Svensson et al. [2]. As was previously mentioned in this section, we aim to more precisely close this gap between the use of Kalman prediction techniques and optimal alarm systems in this article. However, this article is not meant to serve as a cure-all to the ongoing debate, but rather as a participation in this discussion from a different theoretical angle, infused by a segment of the literature that has been largely overlooked. Furthermore, it is motivated by practical examples whose anomalies can be described from multiple variants within a class of level-crossing events in lieu of only one.

II. Motivation for Target Applications

Traditionally, examples of failures using anomaly detection techniques can be characterized by a level-crossing of a critical level, L, that is assumed to have a fixed, static value. The level is exceeded by some critical parameter than can be represented by a dynamic process, which can often be modeled as a zero-mean linear dynamic system driven by Gaussian noise. Most of the theory that follows is based upon this standard representation of the anomaly detection problem.

A. Thermal Comfort Application

For the example currently under discussion based on thermal sensation complaints, the critical value L is not fixed. In fact, it varies with time, and there are two of these stochastic critical levels: one for hot complaints, and another for cold. These levels represent the temperature at which a group of occupants in a zone would complain if too hot or cold. They are somewhat artificial, because such temperatures cannot be measured continuously. However, when complaints do occur, the temperatures can be measured and stored in a maintenance management database. Therefore the statistics of these levels can be computed from this database repository, as described by Federspiel [14], and used to generate a model whose output represents the complaint levels of interest.

The two processes to be used for this example are one of the two stochastic critical levels (i.e., the hot complaint level) and its interaction with the controlled process of interest (i.e., the building or space/zone temperature). In order to transform this problem easily into one that fits the paradigm of a fixed, static threshold, we simply take the difference between the stochastic critical level and the controlled process, implying that $L = 0$. Since there are two stochastic critical levels, both hot and cold, with differing descriptive statistics, these alarm systems will need to be designed independently and implemented in parallel. Only one will be presented in the subsequent sections for illustrative purposes. The idea of the critical level itself being modeled as a stationary Gaussian process was also recognized by Svensson as a potential candidate for study in the context of optimal alarm systems [1], p.93.

In buildings, time-of-day complaint rates and energy usage fluctuate in a predictable manner. Therefore, it is common to expect a peak in the complaint rate during the morning, called the "arrival complaint period." The arrival complaint phenomenon was hypothesized by Federspiel et al. [15] as the result of a naturally high metabolic rate of building occupants during this period. Hence we can look at breaking down the periods of interest into two distinct timeframes, described below.

1) Arrival and operating complaints: Prior to the start of the beginning of the day (eg. 8 am), we want to predict an arrival complaint, and all remaining operating complaints for a sliding window of time of fixed length. An arrival complaint has no restriction on happening at a particular time (i.e., late arrivals are allowed).

2) Operating complaints only: Following the start of the day (i.e., conceivably after the first arrival complaint), we want to predict all operating complaints any point after the beginning of the day, for a sliding window of time of fixed length.

These timeframes each correspond to different level-upcrossing events, to be discussed for this specific application in more detail in the subsequent section. However, regardless of application, upcrossings, downcrossings, and exceedances are defined as follows:

1) Exceedance: A one-dimensional level-crossing event, $\{x_k > L\}$, where L is some critical threshold level exceeded by a process whose value at time k is x_k.

2) Upcrossing: A two-dimensional level crossing event, $\{x_k < L,\ x_{k+1} > L\}$.

3) Downcrossing: A two-dimensional level crossing event $\{x_k > L,\ x_{k+1} < L\}$.

The ability to predict the average thermal comfort of a group of building occupants within a zone during either of the two timeframes listed previously can aid abundantly in developing optimal thermostat setting strategies. Automation of some of the critical decisions that facility mangers often do not have adequate time to attend to within the building operations domain can potentially help to save significantly in operating costs.

B. Spacecraft Propulsion System Anomaly Detection Application

The primary parameter of interest for this study that is available and measured for spacecraft propulsion systems is the control system error, or the difference between the commanded and actual throttle. Not only is this practically appealing due to the the fact that there are often hard limits set on the control system error, but the novel methods described in this paper apply quite cleanly to this parameter. This is evident due to the zero mean of the control system error during non-transient operation, and qualitatively Gaussian characteristics. Furthermore, the control systems were most likely designed with disturbance rejection in mind. Therefore, any anomalous excursions away from the reference value not explained by

transients is cause for alarm. As such, the design of a robust detection algorithm and subsequent diagnostic investigation are of paramount importance in the implementation and deployment of such an alarm system.

A very practical solution to the detection problem is the use of hard thresholds, also commonly known as "redlines." These limits on throttle control system error act as a basic, yet very effective measure of implementing anomaly detection-based alarm systems. A mixture ratio control system used aboard a spacecraft propulsion system may also benefit from the application of a similar detection algorithm. However, redlines are used for a variety of other parameters not used for control, and the zero-mean linear dynamic system driven by Gaussian noise modeling paradigm may break down accordingly. Further investigation of the modeling variants for different spacecraft propulsion system parameters can be found in [9]. In this article, one of the primary investigative themes for this application is to compare the redline detection method to others, including the novel one introduced in this paper.

Unlike the thermal comfort application, the level-crossing problem can't easily be transformed into upcrossings of a level $L = 0$ for either hot or cold complaints. This example requires a little more complexity due to the nature of the parameter of interest. However, there is no need to transform the level-upcrossing problem to be commensurate with a level of $L = 0$. Rather, because the magnitude of the control system error is of interest, the absolute value of the controlled process exceeding a non-zero level, L, becomes the application-specific problem of interest. Here again, there are two critical levels of interest, one above the controlled process and one below. In this case, neither is a stochastic critical level, and both static, fixed thresholds are symmetric about zero and the negative of each other. As such, there is no need to design independent alarm systems that are implemented in parallel. Rather, a single alarm system can be designed to predict all necessary crossing events.

III. General Approach

In certain cases, specifically for the type of level-upcrossing events relevant to the thermal comfort application, the complicated multi-dimensional level-crossing event can be approximated using a variety of methods. The theoretical derivations and comparisons of these different approximations are left out of this article for clarity of presentation. We refer readers to [3] for more details. While cumbersome to present here, a thorough understanding of these details is necessary to fully appreciate the notation used in subsequent sections.

Therefore, we introduce the basic notation as a convenience to the reader in the following section. The main notational emphasis will be on characterization of the crossing events in order to determine alarm regions resulting from use of the likelihood ratio resulting in the conditional inequality: $P(C|\mathcal{D}) \geq P_b$. This basically says "give alarm when the conditional probability of the event, C, exceeds the level P_b." Here, $\mathcal{D}$ represents data being conditioned on, and P_b represents some optimally chosen border or threshold probability with respect to a relevant alarm system metric. It is necessary to find the alarm regions in order to design the alarm system.

Alarm design requires computation of the metrics that characterize the tradeoff that all such systems contend with. This tradeoff represents the balance of false alarms versus missed detections. There are several alarm system metrics to choose from, among them are the ROC curve, percent accuracy, precision-recall curve, and Type I/II error probabilities, all for various border probabilities (P_b). Type I/II error probabilities are the false positive rate and probability of missed detection, respectively.

The false alarm probability can also be a useful candidate for an alarm system metric as an alternative to the Type I or false positive rate. Technically, if the false alarm probability is defined as the ratio of the number of false alarms to the number of alarms rather than the number of negatives, it is not a Type I probability. Here we define Type I probability as the false positive rate, or the ratio of the number of false alarms to the number of negatives.

In alarm design, we want to find the value of P_b that provides the best tradeoff between Type I and Type II errors, or one of the other alarm system metrics. For the first example involving the thermal comfort application, most of the analysis was performed by using Type II error and false alarm probabilities (although incorrectly annotated as Type I/II in [3]). However, for the spacecraft propulsion system application, the ROC curve will be the metric used for comparison of algorithms. The reason for using the ROC curve is that it is more reliable in the face of uneven examples of nominal vs. anomalous behavior, as described in [16].

It is not possible to obtain the exact alarm system metrics analytically, or even by means of numerical integration for the complicated multi-dimensional events we will present here. As an alternative, we may perform simulations to obtain an *estimate* of the exact alarm system metrics. These simulation-based statistics have well known estimation error properties. They are obtained by running a Kalman predictor, and counting the number of correct/false alarms and missed detections until their relative frequencies converge to limiting probability values. However, with the aid of some approximations, we can perform numerical integrations of complex integrals, and can avoid these otherwise often very time and computationally intensive simulation runs.

In some cases the number of terms required to compute the relevant probability-based alarm system metrics scale exponentially with the number of time steps under consideration. This is particularly true for the types of level crossing events characterizing complaints for the thermal comfort application. As such, approximations are developed to reduce the resulting computational burden. One approximation, the multi-dimensional approximation, can characterize either of the two null hypotheses to be introduced and tested in the subsequent sections. Both correspond to the breakdown of the periods of interest into two distinct timeframes presented previously. Although this is the approximation in which the number of terms scale exponentially with the number of steps in the time interval, it may also be used to test null hypotheses which quantify any combination of upcrossings, downcrossings, or exceedances.

The multivariate probability computations that result from

the theory presented later are performed by using Genz's algorithm [17]. This algorithm is based upon a robust technique designed to be used for integrations in multiple dimensions. Traditionally, this code is more effective and computationally efficient for higher-dimensional integrations, but can be used just as well for lower-dimensional ones. As such, it helps to mitigate the exponentially scaling computational burden. Furthermore, it provides a method of computing integrals necessary for the design of optimal alarm systems, and also other failure detection algorithms such as the one most often used by Kerr [18], who specifically cites issues with the computation of these types of integrals.

One might question the merit of using such an unavoidably costly and potentially computationally intensive technique to design an alarm system, when a simpler one might do. There are several different types of alarm systems, ranging in frequency of use and expense. At one end, we have alarm systems that require little economic investment other than the accumulation of man-hours of heuristic knowledge. These systems, although quite inexpensive and often lacking in sophistication, tend to be the most ubiquitous in engineering systems. For the most part, they tend to the job that is required of them: to give alarm for prevention of catastrophic events. False alarms would cause loss of production and capital due to system downtime from the inevitable system shutdown as a result of the alarm. However, missed detections may cause damage, complete system destruction or loss of life, where the costs are immeasurable. Since complex engineering systems may encounter events that need to be predicted by these simple alarm systems, more sophisticated alarm systems may be of interest. The following list provides a variety of alarm systems to be compared on an application-specific level, ranging from the most simple one to the more sophisticated ones alluded to previously.

A. "Redlines" or Simple Alarm System

Typically there is no computational design cost for this type of alarm system, but rather the cost lies in the knowledge and experience of the users, i.e., heuristics. The basic idea is that certain thresholds are chosen apriori to provide a window of operation within which a random or controlled process with random components should be constrained. This alarm system will be investigated for the spacecraft propulsion system example only.

B. Predictive Alarm System

An alarm system that uses a predictive method is often called a naïve alarm system [6], [1], [2]. Here, a predicted future process value would trigger an alarm if it exceeds some fixed, pre-selected alarm threshold. However, even though the predictor may be optimal in the least-squares sense, the alarm system would not be optimal in the sense that it triggers the fewest false alarms for a fixed detection probability. This alarm system will be investigated for the spacecraft propulsion system example only.

C. Optimal Alarm Systems

Described before, an optimal alarm system is derived based upon a likelihood ratio criterion via the Neyman-Pearson lemma. The resulting optimal alarm system requires the use of predicted future process values to elicit the fewest possible false alarms for a fixed detection probability. As stated earlier, there are several approximations which may be used as an alternative to designing the optimal alarm system based upon simulation, i.e., the "counting method." These approximations are listed below, and are the ones to be studied in depth in this paper.

1) Two-Dimensional Alarm System: Also called a semi-naïve alarm system in the literature [1], [2], it uses the idea of optimal alarm. However, the two-dimensional alarm region is approximated with asymptotes to "rectangularize" it, making the region of integration much easier to define. This alarm system will be investigated for both applications.

2) Multiple Sub-Interval Alarm System: We may also use the union of disjoint sub-intervals to approximate the exact alarm region. This approximation was again a recommendation of Svensson [1], although not elaborated on in detail, and only meant to capture a single level crossing over a time period. Multiple level crossings over a time period often involve complicated multi-dimensional events. Therefore, aggregating less complicated 2-dimensional alarm regions reduces the computational load and increases mathematical tractability. Each sub-interval can be approximated with asymptotes, again making the regions of integration much easier to define. This alarm system will be investigated for the thermal sensation complaint example only.

3) Multi-Dimensional Alarm System: The exact alarm system metrics cannot feasibly be obtained for complicated multi-dimensional events by means of numerical integration. However, an approximate alarm region of integration can be defined as a tight bound on the exact region via the unions and/or intersections of hyperplanes. In certain cases, this approximation forms a semi-infinite hyper-rectangular region in $\mathbb{R}^n$, where n is the dimension of the space, or number of predictive time steps under consideration. This alarm system will be investigated for both applications.

IV. Theory

A. Kalman Filtering and Prediction

Before attempting to technically characterize the alarm regions of interest or explain the approximation methods, it is necessary to address the basic mathematical paradigms that we'll use. As such, let's assume that a stationary, Gaussian, random process can be characterized in state-space as a typical linear system of the form

$$\begin{aligned} \mathbf{q}_{k+1} &= \mathbf{A}\mathbf{q}_k + \mathbf{w}_k \\ x_k &= \mathbf{C}\mathbf{q}_k + v_k \end{aligned} \tag{1}$$

where $\mathbf{q}_k$ is the unobserved state of the process with measured output x_k. Apriori statistics for the input and measurement noise sequences, $\mathbf{w}_k$ and v_k, also need to be defined. Their covariances are

$$\begin{aligned}\mathbf{Q} &\triangleq E[\mathbf{w}_k\mathbf{w}_k^T]\\ R &\triangleq E[v_k v_k^T]\end{aligned}$$

We also assume $\mathbf{w}_k$ and v_k are zero-mean Gaussian white noise sequences without loss of generality, such that $\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$ and $v_k \sim \mathcal{N}(0, R)$.

Propagation of the *unconditional* covariance matrix is shown below in Eqn. 2.

$$\mathbf{P}_{k+1} = \mathbf{A}\mathbf{P}_k\mathbf{A}^T + \mathbf{Q} \tag{2}$$

where $\mathbf{P}_k \triangleq E[\mathbf{q}_k\mathbf{q}_k^T]$.

The algebraic equivalent to this propagation equation is $\mathbf{P}_{ss}^L = \mathbf{A}\mathbf{P}_{ss}^L\mathbf{A}^T + \mathbf{Q}$, where $\mathbf{P}_{ss}^L \succ 0$ ($\mathbf{P}_{ss}^L$ is positive definite) and it is also the solution to this discrete algebraic Lyapunov equation. The time and measurement update steps of the recursive Kalman filtering equations are shown in Eqns. 3-5. Eqn. 4 represents the Kalman gain.

$$\begin{aligned}\hat{\mathbf{q}}_{k+1|k} &= \mathbf{A}\hat{\mathbf{q}}_{k|k} &(3)\\ \mathbf{F}_{k+1|k} &\triangleq \mathbf{P}_{k+1|k}\mathbf{C}^T(\mathbf{C}\mathbf{P}_{k+1|k}\mathbf{C}^T + R)^{-1} &(4)\\ \hat{\mathbf{q}}_{k+1|k+1} &= \hat{\mathbf{q}}_{k+1|k} + \mathbf{F}_{k+1|k}(x_{k+1} - \mathbf{C}\hat{\mathbf{q}}_{k+1|k}) &(5)\end{aligned}$$

where $\hat{\mathbf{q}}_{k|k} \triangleq E[\mathbf{q}_k|x_0, \ldots, x_k]$.

The counterpart *conditional* covariance propagation time and measurement updates for the Kalman filter are shown in Eqns. 6-7, respectively.

$$\begin{aligned}\mathbf{P}_{k+1|k} &= \mathbf{A}\mathbf{P}_{k|k}\mathbf{A}^T + \mathbf{Q} &(6)\\ \mathbf{P}_{k+1|k+1} &= \mathbf{P}_{k+1|k} - \mathbf{F}_{k+1|k}\mathbf{C}\mathbf{P}_{k+1|k} &(7)\end{aligned}$$

where

$$\mathbf{P}_{k|k} \triangleq E[(\mathbf{q}_k - \hat{\mathbf{q}}_{k|k})(\mathbf{q}_k - \hat{\mathbf{q}}_{k|k})^T|x_0, \ldots, x_k]$$

Combining the two equations 6 and 7, we obtain the following:

$$\mathbf{P}_{k+1|k} = \mathbf{A}\mathbf{P}_{k|k-1}\mathbf{A}^T - \mathbf{A}\mathbf{F}_{k|k-1}\mathbf{C}\mathbf{P}_{k|k-1}\mathbf{A}^T + \mathbf{Q} \tag{8}$$

The stationary version of Eqn. 8 gives us the solution to the discrete algebraic Riccati equation, as follows:

$$\begin{aligned}\mathbf{P}_{ss}^R &= \mathbf{A}\mathbf{P}_{ss}^R\mathbf{A}^T - \mathbf{A}\mathbf{F}_{ss}\mathbf{C}\mathbf{P}_{ss}^R\mathbf{A}^T + \mathbf{Q} &(9)\\ \mathbf{F}_{ss} &\triangleq \mathbf{P}_{ss}^R\mathbf{C}^T(\mathbf{C}\mathbf{P}_{ss}^R\mathbf{C}^T + R)^{-1} &(10)\end{aligned}$$

where $\mathbf{F}_{ss}$ represents the steady-state Kalman gain. However, we're interested in the updated aposteriori steady-state covariance matrix, which is the stationary version of Eqn. 7 given by:

$$\hat{\mathbf{P}}_{ss}^R = \mathbf{P}_{ss}^R - \mathbf{P}_{ss}^R\mathbf{C}^T(\mathbf{C}\mathbf{P}_{ss}^R\mathbf{C}^T + R)^{-1}\mathbf{C}\mathbf{P}_{ss}^R \tag{11}$$

Because we're primarily concerned with *prediction*, we will need to compute variances of the form

$$V_{k+i|k} \triangleq \text{Var}(x_{k+i}|x_0, \ldots, x_k)$$

and covariances of the form

$$\text{cov}(x_{k+i}, x_{k+j}|x_0, \ldots, x_k)$$

where i, j are prediction window indices for future process values. It can be shown that $V_{k+i|k}$ and its covariance counterparts are functions of $\mathbf{P}_{k|k}$ and $\mathbf{P}_k$, and therefore $\hat{\mathbf{P}}_{ss}^R$, and $\mathbf{P}_{ss}^L$, respectively. As a result, they can be expressed as being independent of the time index k, although they cannot be expressed as being independent of the prediction window indices i, j. Similarly, we will need to compute the predicted future process value $\hat{x}_{k+i|k}$.

$$\begin{aligned}\hat{x}_{k+i|k} &\triangleq E[x_{k+i}|x_0, \ldots, x_k] &(12)\\ &= \mathbf{C}E[\mathbf{q}_{k+i}|x_0, \ldots, x_k] &(13)\\ &= \mathbf{C}\hat{\mathbf{q}}_{k+i|k} &(14)\end{aligned}$$

Obviously, $\hat{x}_{k+i|k}$ can be expressed as a function of $\hat{\mathbf{q}}_{k|k}$, but unlike $V_{k+i|k}$, it will fluctuate as new measurements are made. This is apparent due to Eqn. 5, which is directly dependent upon measurements x_{k+1}.

B. Alarm Regions for Crossing Events and their Approximations

1) Thermal Comfort Level Crossings: We first present the conditions for alarm based upon the thermal sensation complaint application. The null hypothesis, $\mathcal{H}_0$, shown in Eqn. 15 is for at least one complaint during normal building operating hours, where m refers to the number of steps in the prediction window. The $m+1$ dimensional event region is given by $\Omega_{C_{exact}}$.

$$\begin{aligned}\mathcal{H}_0 &: (\mathbb{X} \in \Omega_{C_{exact}} \subset \mathbb{R}^{m+1}) &(15)\\ \mathbb{X} &= \begin{bmatrix} x_k \\ \vdots \\ x_{k+m} \end{bmatrix}\\ C_{exact} &= \{\mathbb{X} \in \Omega_{C_{exact}} \subset \mathbb{R}^{m+1}\}\\ &= \neg\{x_k < L, \ldots, x_{k+m} < L\} \setminus \{x_k > L\}\end{aligned}$$

In Eqn. 15, $\neg$ is the logical equivalent of *not*, and the event given by $\{x_k < L, \ldots, x_{k+m} < L\}$ refers to the fact that there are no level crossings of $L = 0$ (complaints) from time k to $k+m$, or $x_j < L, \forall j \in \{k, \ldots, k+m\}$. The $\setminus$ notation denotes set subtraction of the event defined by $\{x_k > L\}$, which corresponds to arrival complaints.

In this case, the condition for alarm leads to the inequality shown in Eqn. 16, via the Neyman-Pearson lemma [1], [2]. This inequality and Eqn. 15 are meant to characterize the "operating complaints only" scenario as accurately as possible. In Eqn. 15, operating complaints can be defined with any sequence of process values above or below the critical threshold, as long as **all** of the values don't lie above the threshold, and

the first process value is not above the threshold (to keep from counting arrival complaints).

$$P(x_k < L|\mathcal{D}) - P(x_k < L, \ldots, x_{k+m} < L|\mathcal{D}) \geq P_b \quad (16)$$

where: $\mathcal{D} = \{x_0, \ldots, x_k\}$. The corresponding exact alarm condition can be partitioned as follows:

$$\begin{aligned} A_{exact} &\triangleq \{\hat{\mathbb{X}} \in \Omega_{A_{exact}} \subset \mathbb{R}^{m+1}\} \\ &= \{\hat{\mathbb{X}} : P(C_{exact}|\mathcal{D}) \geq P_b\} \\ &= \{\hat{\mathbb{X}} : P(x_k < L|\mathcal{D}) \\ &\quad -P(x_k < L, \ldots, x_{k+m} < L|\mathcal{D}) \geq P_b\} \end{aligned}$$

where:

$$\mathbb{X} = \begin{bmatrix} x_k \\ x_{k+1} \\ \vdots \\ x_{k+m} \end{bmatrix}, \ \hat{\mathbb{X}} = E[\mathbb{X}|\mathcal{D}] = \begin{bmatrix} \hat{x}_{k|k} \\ \hat{x}_{k+1|k} \\ \vdots \\ \hat{x}_{k+m|k} \end{bmatrix}$$

Hence it is easy to write the formulae for correct alarms/detections:

$$\text{Let } \mathbf{x} = \begin{bmatrix} \mathbb{X} \\ \hat{\mathbb{X}} \end{bmatrix} \quad (17)$$

Correct Alarm:

$$\begin{aligned} P(C_{exact}|A_{exact}) &= \frac{P(C_{exact}, A_{exact})}{P(A_{exact})} \quad (18) \\ &= \frac{\int_{\Omega_{C_{exact}}} \int_{\Omega_{A_{exact}}} \mathcal{N}(\mathbf{x}; \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}}) \, d\mathbf{x}}{\int_{\Omega_{A_{exact}}} \mathcal{N}(\hat{\mathbb{X}}; \mu_{\hat{\mathbb{X}}}, \Sigma_{\hat{\mathbb{X}}}) \, d\hat{\mathbb{X}}} \end{aligned}$$

Correct Detection:

$$\begin{aligned} P(A_{exact}|C_{exact}) &= \frac{P(C_{exact}, A_{exact})}{P(C_{exact})} \quad (19) \\ &= \frac{\int_{\Omega_{C_{exact}}} \int_{\Omega_{A_{exact}}} \mathcal{N}(\mathbf{x}; \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}}) \, d\mathbf{x}}{\int_{\Omega_{C_{exact}}} \mathcal{N}(\mathbb{X}; \mu_{\mathbb{X}}, \Sigma_{\mathbb{X}}) \, d\mathbb{X}} \end{aligned}$$

These probabilities are necessary in order to compute the required alarm system metrics. The probability of correct alarm measures the ability not to generate false alarms (purity), and the probability of correct detection measures that ability not to miss any critical anomalies (completeness). We know that computing these integrals, specifically the alarm region, A_{exact}, is an intractable problem. Therefore we must use a simulation to obtain estimates of such probabilities. But instead of using this "counting" method via simulation, we can compute their tractable and much less computationally intensive approximations.

The first approximation for the exact alarm region corresponding to operating complaints only and given by $A_{exact} = \{\hat{\mathbb{X}} : P(x_k < L|\mathcal{D}) - P(x_k < L, \ldots, x_{k+m} < L|\mathcal{D}) \geq P_b\}$ is a single interval two-dimensional approximation. In this case we only consider two time slices, $\{i = 0, \ i = m\}$, which span the entire time interval being considered. The process value, x_{k+i} is below the critical threshold at the very beginning of the interval, x_k, and above it at the very end, x_{k+m}. The exact null hypothesis captures at least one complaint or upcrossing, where the process itself is not restricted to being above or below the critical threshold at the end of the interval. Therefore, this approximation will miss around half of the upcrossings/complaints caught by the exact condition. However, using it will greatly reduce computation time, and the answer can be achieved via numerical integration, without the use of simulation.

The upcrossing approximation is shown in Eqns. 20-24. By using it, not only do we reduce the dimension of the alarm region from $m+1$ to 2, shown in the first step via Eqn. 20, but a "rectangularized" two-dimensional approximation is used for further ease of computation. Recall that a two-dimensional alarm system is one that actually uses the idea of optimal alarm. The exact $2D$ alarm region is approximated with asymptotes, making the revised region of integration much easier to parameterize. The "rectangularization" is apparent in Fig. 1.

To find the asymptote corresponding to $\hat{x}_{k|k}$, we derive the limiting distribution as the remaining dimension is marginalized by taking $\lim_{\hat{x}_{k+m|k} \to \infty}$ of Eqn. 20, yielding Eqn. 21,

$$\begin{aligned} &P(x_k < L, x_{k+m} > L|x_0, \ldots, x_k) \geq P_b \quad (20) \\ &\lim_{\hat{x}_{k+m|k} \to \infty} P(x_k < L, x_{k+m} > L|x_0, \ldots, x_k) = \\ &\qquad P(x_k < L|x_0, \ldots, x_k) \end{aligned}$$

$$\begin{aligned} &P(x_k < L|x_0, \ldots, x_k) \geq P_b \\ &\Updownarrow \\ &\hat{x}_{k|k} < L - \sqrt{V_{k|k}}\Phi^{-1}(P_b) \quad (21) \end{aligned}$$

where $\Phi^{-1}(\cdot)$ represents the inverse cumulative normal standard distribution function. Similarly, to find the asymptote corresponding to $\hat{x}_{k+m|k}$, we derive the limiting distribution as the remaining dimension is marginalized by taking $\lim_{\hat{x}_{k|k} \to -\infty}$ of Eqn. 22, yielding Eqn. 23.

$$\begin{aligned} &P(x_k < L, x_{k+m} > L|x_0, \ldots, x_k) \geq P_b \quad (22) \\ &\lim_{\hat{x}_{k|k} \to -\infty} P(x_k < L, x_{k+m} > L|x_0, \ldots, x_k) = \\ &\qquad P(x_{k+m} > L|x_0, \ldots, x_k) \end{aligned}$$

$$\begin{aligned} &P(x_{k+m} > L|x_0, \ldots, x_k) \geq P_b \\ &\Updownarrow \\ &\hat{x}_{k+m|k} > L + \sqrt{V_{k+m|k}}\Phi^{-1}(P_b) \quad (23) \end{aligned}$$

where:

$$\begin{aligned} V_{k+i|k} &\triangleq \text{Var}(x_{k+i}|x_0, \ldots, x_k) \quad (24) \\ &= \mathbf{C}(\mathbf{A}^i(\hat{\mathbf{P}}^R_{ss} - \mathbf{P}^L_{ss})(\mathbf{A}^T)^i + \mathbf{P}^L_{ss})\mathbf{C}^T + R \end{aligned}$$

The alarm region can therefore be approximated by the two-dimensional intersection of the two inequalities represented in Eqns. 21 and 23.

Notice that there are several two-dimensional alarm regions shown in Fig. 1, for values of P_b ranging from 0.1 to 0.9, in

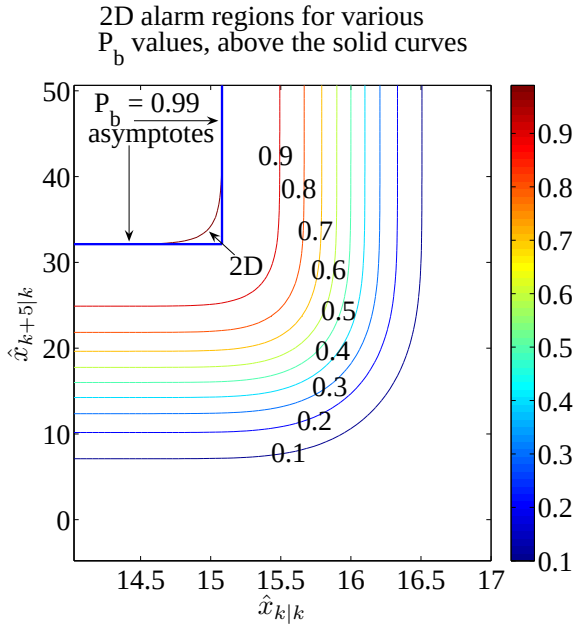

Fig. 1. Approximation to the 2D Alarm Region for $P_b = 0.99$

gradations of 0.1, all for $m = 5$. Each of the regions is convex, above and to the left of which is considered the alarm region.[1] However, performing an integration over this two-dimensional alarm region may require storing the contour points along its border, or other approximation methods that consume quite a bit more compute cycles than the respective "rectangularized" $2D$ alarm region. For the specific case of $P_b = 0.99$, we can see the asymptotes in Fig. 1 that define the approximation, which bound the two-dimensional alarm region. Integration over this region is much easier, and the magnitude of the error introduced by this additional approximation will not be on par with the approximation of the "$m+1$"-dimensional exact region with a reduced 2-dimensional region.

To improve upon the approximation introduced in the previous section, and "catch" more of the cases missed by the single two-dimensional interval, we can split the interval into N_s disjoint two-dimensional subintervals, and construct alarm systems for each subinterval. Here we'd like to compute relevant aggregate alarm system design probabilities for the entire interval in question, by taking the *union* of the alarm systems corresponding to each subinterval. Practically, this means that if any one of the N_s sub-interval based alarm systems sound, then the alarm system as a whole sounds. Obtaining the approximate two-dimensional alarm regions for each subinterval is easy, and is based upon the same logic in Eqns. 20-24. In order to determine the aggregate alarm system design probabilities, we use Eqns. 29-32. But first we need to define the alarm sub-interval approximation provided in Eqns. 25-28.

[1]These sample alarm regions were generated by using a very simple example (c.f. Process 1 from Svensson et. al. [2]) solely for illustrative purposes at this point. None of the alarm regions are based on models generated from experimental data.

First, recall:

$$C_{exact} = \neg\{x_k < L, \ldots, x_{k+m} < L\} \setminus \{x_k > L\} \quad (25)$$

The resulting alarm sub-interval approximations are as follows:

$$A_i = \{\hat{x}_{k+s_i|k} < \mathcal{X}_{s_i}, \hat{x}_{k+s_{i+1}|k} > \mathcal{Y}_{s_{i+1}}\} \quad (26)$$

$$\mathcal{X}_{s_i} = L - \sqrt{V_{k+s_i|k}}\Phi^{-1}(P_b) \quad (27)$$

$$\mathcal{Y}_{s_{i+1}} = L + \sqrt{V_{k+s_{i+1}|k}}\Phi^{-1}(P_b) \quad (28)$$

where the extremes of the interval $0 \leq s_i, s_{i+1} \leq m$ correspond to endpoints of all possible subintervals, given by indices s_i and s_{i+1}. The approximate alarm region can therefore be written as $A_{approx} \triangleq \bigcup_{i=1}^{N_s} A_i$. Formulae for correct/false alarms and correct/missed detections can be developed, as shown in Eqns. 29-32.

Correct Alarm:

$$P(C_{exact}|A_{approx}) = \frac{P(C_{exact}, A_{approx})}{P(A_{approx})} \quad (29)$$

Correct Detection:

$$P(A_{approx}|C_{exact}) = \frac{P(C_{exact}, A_{approx})}{P(C_{exact})} \quad (30)$$

False Alarm:

$$P(C'_{exact}|A_{approx}) = \frac{P(C'_{exact}, A_{approx})}{P(A_{approx})} = 1 - P(C_{exact}|A_{approx}) \quad (31)$$

Missed Detection:

$$P(A'_{approx}|C_{exact}) = \frac{P(C_{exact}, A'_{approx})}{P(C_{exact})} = 1 - P(A_{approx}|C_{exact}) \quad (32)$$

In order to compute the deceivingly simple-looking formulae in Eqns. 29-32, we need more detailed equations for $P(A_{approx})$, $P(C_{exact}, A_{approx})$, and $P(C_{exact})$, which are provided in [3]. Recall that the central idea of the multiple subinterval approximation method is that less complicated 2-dimensional alarm regions are aggregated, thus reducing the computational load and increasing mathematical tractability. Each sub-interval can be approximated with asymptotes, again making the regions of integration much easier to define. The case used to illustrate the approximation above was for a fixed critical threshold, considering operating complaints only.

If we were to have considered both arrival and operating complaints, a different alarm region would have resulted. As such, it is worthwhile to consider the fact that arrival complaints are defined as "exceedances," defined earlier. Because exceedances are defined in single time slices, we don't know about the arrival complaint until the exceedance terminates in some subsequent time slice. Therefore, the arrival complaint may best be determined with knowledge of a downcrossing, as opposed to an upcrossing.

The final approximation method is meant to provide the best possible approximation to the alarm region for the thermal

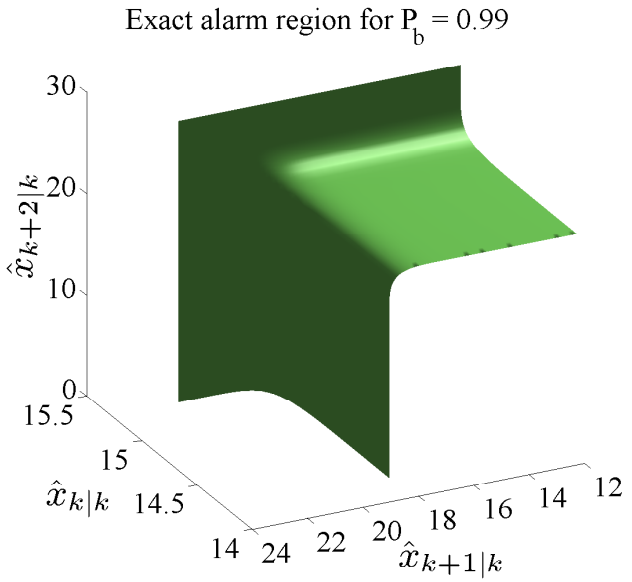

Fig. 2. Exact Alarm Region for $m = 3$

sensation complaint application, given by $A_{exact} = \{\hat{\mathbb{X}} : P(x_k < L|\mathcal{D}) - P(x_k < L, x_{k+1} < L, \ldots, x_{k+m} < L|\mathcal{D}) \geq P_b\}$. We know that the alarm region does not serve as a well-defined region of integration. Therefore, the *exact* alarm system metrics cannot feasibly be obtained for complicated multi-dimensional events by means of numerical integration. To illustrate this fact, Fig. 2 shows the exact alarm region, for operating complaints only, when $m = 3$ and the multivariate Gaussian integrand is 4-dimensional. The alarm region is *above* and to the left of the surface shown.[2]

It is apparent from the figure that the surface of the exact alarm region boundary is quite complex and does not serve as a feasible, parameterizable integration region. However, an approximate volume can be defined as a tight bound on the exact region via the unions and/or intersections of planes. This approximation forms a semi-infinite hyper-rectangular region in multi-dimensional space. For operating complaints only, the region of integration, A_{approx}, has the following representation:

[2]This is also a sample alarm region generated by the same example as in Fig. 1. It is provided solely for illustrative purposes at this point and not based on models generated from experimental data.

$$
\begin{aligned}
A_{exact} &\triangleq \{\hat{\mathbb{X}} \in \Omega_{A_{exact}} \subset \mathbb{R}^{m+1}\} \\
&= \{\hat{\mathbb{X}} : P(x_k < L|\mathcal{D}) - \\
&\quad P(x_k < L, \ldots, x_{k+m} < L|\mathcal{D}) \geq P_b\} \\
&\Updownarrow \text{ Multi-dimensional approximation} \\
A_{approx} &\triangleq \{\hat{\mathbb{X}} \in \Omega_{A_{approx}} \subset \mathbb{R}^{m+1}\} \\
&= \left\{\hat{\mathbb{X}} : A_0 \bigcap \left[\bigcup_{i=1}^{m} A_i\right]\right\} \qquad (33) \\
&\qquad\qquad \overbrace{\qquad\qquad}^{\mathcal{X}_0} \\
\text{where } A_0 &\triangleq \hat{x}_{k|k} \leq L - \sqrt{V_{k|k}}\Phi^{-1}(P_b) \\
A_i &\triangleq \hat{x}_{k+i|k} \geq \underbrace{L + \sqrt{V_{k+i|k}}\Phi^{-1}(P_b)}_{\mathcal{Y}_i} \\
\forall i &\in 1, \ldots, m
\end{aligned}
$$

The multi-dimensional approximation shown in the chain of logic above can be further elaborated on. For all asymptotes corresponding to $\hat{x}_{k+i|k}, \forall i > 0$, we have:

$$\lim_{\hat{\mathbb{X}}\setminus\hat{x}_{k+i|k}\to-\infty} P(x_k < L|\mathcal{D}) - P(x_k < L, \ldots, x_{k+m} < L|\mathcal{D}) =$$

$$1 - P(x_{k+i} < L|x_0, \ldots, x_k)$$

and

$$1 - P(x_{k+i} < L|x_0, \ldots, x_k) \geq P_b, \forall i > 0$$

$$\Updownarrow$$

$$\bigcup_{i=1}^{m} \mathcal{Y}_i \qquad (34)$$

Furthermore, for the remaining asymptote corresponding to $\hat{x}_{k|k}$, we have:

$$\lim_{\hat{\mathbb{X}}\setminus\hat{x}_{k|k}\to\infty} P(x_k < L|\mathcal{D}) - P(x_k < L, \ldots, x_{k+m} < L|\mathcal{D}) =$$

$$\lim_{\hat{\mathbb{X}}\setminus\hat{x}_{k|k}\to\infty} P(x_k < L, \underbrace{x_{k+1} > L, \ldots, x_{k+m} > L}_{x_{k+i}>L, \forall i\in[1,m]}|\mathcal{D}) + \ldots$$

$$\lim_{\hat{\mathbb{X}}\setminus\hat{x}_{k|k}\to\infty} \sum_j P(x_k < L, \mathcal{T}_j|\mathcal{D})$$

where

$$\mathcal{T}_j \triangleq \left\{\bigcap_{i=1}^{m} (x_{k+i} > L \vee_j x_{k+i} < L) \mid \exists i : x_{k+i} < L\right\}$$

$$\lim_{\hat{\mathbb{X}}\setminus\hat{x}_{k|k}\to\infty} P(x_k < L, \underbrace{x_{k+1} > L, \ldots, x_{k+m} > L}_{x_{k+i}>L, \forall i\in[1,m]}|\mathcal{D}) =$$

$$P(x_k < L|x_0, \ldots, x_k)$$

and due to the definition of the set $\mathcal{T}_j$, if $\exists i : x_{k+i} < L$, then

$$\lim_{\hat{\mathbb{X}}\backslash\hat{x}_{k|k}\to\infty}\sum_j P(x_k < L, \mathcal{T}_j|\mathcal{D}) = 0$$

Therefore

$$\lim_{\hat{\mathbb{X}}\backslash\hat{x}_{k|k}\to\infty} P(x_k < L|\mathcal{D}) - P(x_k < L, \ldots, x_{k+m} < L|\mathcal{D}) =$$

$$P(x_k < L|x_0, \ldots, x_k)$$

and

$$P(x_k < L|x_0, \ldots, x_k) \geq P_b$$

$$\Updownarrow$$

$$\hat{x}_{k|k} \leq L - \sqrt{V_{k|k}}\Phi^{-1}(P_b) \tag{35}$$

Taking the intersection of Eqns. 34 and 35 yields the approximation shown in Eqn. 33. A similar procedure can be used to derive the multi-dimensional approximation to the alarm region for both arrival and operating complaints. In this case, the null hypothesis and resulting level-crossing event changes, as shown below in Eqn. 36.

$$\begin{aligned}
\mathcal{H}_0 &: (\mathbb{X} \in \Omega_{C_{exact}} \subset \mathbb{R}^{m+1}) \qquad (36)\\
\mathbb{X} &= \begin{bmatrix} x_{k+d} \\ \vdots \\ x_{k+d+m} \end{bmatrix}\\
C_{exact} &= \{\mathbb{X} \in \Omega_{C_{exact}} \subset \mathbb{R}^{m+1}\}\\
&= \neg\{x_{k+d} < L, \ldots, x_{k+d+m} < L\}
\end{aligned}$$

The resulting the region of integration, A_{approx}, has the following representation:

$$\begin{aligned}
A_{exact} &\triangleq \{\hat{\mathbb{X}} \in \Omega_{A_{exact}} \subset \mathbb{R}^{m+1}\}\\
&= \{\hat{\mathbb{X}} : 1 - P(x_{k+d} < L, \ldots,\\
&\quad x_{k+d+m} < L|\mathcal{D}) \geq P_b\}\\
&\Updownarrow \text{ Multi-dimensional approximation}\\
A_{approx} &\triangleq \{\hat{\mathbb{X}} \in \Omega_{A_{approx}} \subset \mathbb{R}^{m+1}\}\\
&= \left\{\hat{\mathbb{X}} : \bigcup_{i=0}^{m} A_i\right\} \qquad (37)\\
&= \left\{\hat{\mathbb{X}} : \bigcup_{i=0}^{m} \hat{x}_{k+d+i|k} \geq \mathcal{Y}_{d+i}\right\}
\end{aligned}$$

Eqn. 37 gives the approximation for the candidate region of integration for both arrival and operating complaints, and can be derived for all asymptotes corresponding to $\hat{x}_{k+d+i|k}, \forall i \geq 0$, as follows:

$$\lim_{\hat{\mathbb{X}}\backslash\hat{x}_{k+d+i|k}\to-\infty} 1 - P(x_{k+d} < L, \ldots, x_{k+d+m} < L|\mathcal{D}) =$$

$$1 - P(x_{k+d+i} < L|x_0, \ldots, x_k)$$

and

$$1 - P(x_{k+d+i} < L|x_0, \ldots, x_k) \geq P_b, \forall i \geq d$$

$$\Updownarrow$$

$$\bigcup_{i=0}^{m} \hat{x}_{k+d+i|k} \geq \underbrace{L + \sqrt{V_{k+d+i|k}}\Phi^{-1}(P_b)}_{\mathcal{Y}_{d+i}}$$

Note that the future predicted values do not begin until d steps out, in order to allow for a finite window of prediction prior to the beginning of the day. This is distinct from the m steps corresponding to the prediction window during normal building operating hours. The formulae for correct/false alarms and correct/missed detections for both cases are similar to the ones shown in Eqns. 29-32. Again, the computational details of $P(A_{approx})$, $P(C_{exact}, A_{approx})$, and $P(C_{exact})$ are omitted here for the sake of clarity, and are provided in [3].

2) Spacecraft Propulsion System Anomaly Detection Application: As with thermal sensation complaints, anomalies that occur within a spacecraft propulsion system may not have a direct operational mapping to any one type of level-crossing event. Therefore, we will provide a comprehensive review of many different examples within the class of level-crossing events having a fixed threshold. The examples are listed with detailed explanations, all of which use a prediction window denoted by d.

a) Up/downcrossing event spanning an interval: This event is very similar to the single interval two-dimensional approximation described previously for thermal sensation complaints. Recall that two time slices spanning the entire time interval $\{|x_k| < L, |x_{k+d}| > L\}$ are considered. The main difference is that the absolute value of the process is considered, mimicking an envelope detection problem. The logic for this stems from the fact that the control system error is the primary parameter of interest, which can be either positive or negative. Therefore, the process value is within the interval $[-L, L]$ at the very beginning of the interval, and outside of it at the very end. The probability $P(C_{exact})$ is shown in Eqn. 38, and is the same regardless of the alarm system used.

$$\begin{aligned}
P(C_{exact}) &= P(|x_k| < L, |x_{k+d}| > L) \qquad (38)\\
&= \int_{-\infty}^{-L}\int_{-L}^{L} \mathcal{N}(\mathbf{x}; \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})\ d\mathbf{x} + \ldots\\
&\quad \int_{L}^{\infty}\int_{-L}^{L} \mathcal{N}(\mathbf{x}; \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})\ d\mathbf{x}\\
\mathbf{x} &= \begin{bmatrix} x_k \\ x_{k+d} \end{bmatrix}, \mu_{\mathbf{x}} = \begin{bmatrix} \mathbf{C}\mu_{\mathbf{q}} \\ \mathbf{C}\mu_{\mathbf{q}} \end{bmatrix} \in \mathbb{R}^2\\
\Sigma_{\mathbf{x}} &= \begin{bmatrix} \mathbf{C}\mathbf{P}_{ss}^{L}\mathbf{C}^T + R & \mathbf{C}\mathbf{P}_{ss}^{L}(\mathbf{A}^T)^d\mathbf{C}^T \\ \mathbf{C}\mathbf{A}^d\mathbf{P}_{ss}^{L}\mathbf{C}^T & \mathbf{C}\mathbf{P}_{ss}^{L}\mathbf{C}^T + R \end{bmatrix}\\
\mu_{\mathbf{q}} &\triangleq E[\mathbf{q}_k]
\end{aligned}$$

b) End of interval up/downcrossing event: This event is similar to an up/downcrossing event spanning an interval, except that the two time slices are adjacent and the event occurs at the very end of interval, shown mathematically in Eqn. 39.

$$C_{exact} \triangleq \{|x_{k+d}| < L, |x_{k+d+1}| > L\} \tag{39}$$

The probability $P(C_{exact})$ is shown in Eqn. 40, and is also the same regardless of the alarm system used.

$$
\begin{aligned}
P(C_{exact}) &= P(|x_{k+d}| < L, |x_{k+d+1}| > L) \qquad (40)\\
&= \int_{-\infty}^{-L}\int_{-L}^{L} \mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})\ d\mathbf{x} + \ldots\\
&\quad \int_{L}^{\infty}\int_{-L}^{L} \mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})\ d\mathbf{x}\\
\mathbf{x} &= \begin{bmatrix} x_{k+d} \\ x_{k+d+1} \end{bmatrix}\\
\Sigma_{\mathbf{x}} &= \begin{bmatrix} \mathbf{C}\mathbf{P}_{ss}^{L}\mathbf{C}^T + R & \mathbf{C}\mathbf{X}_{ss}\mathbf{C}^T \\ \mathbf{C}\mathbf{X}_{ss}\mathbf{C}^T & \mathbf{C}\mathbf{P}_{ss}^{L}\mathbf{C}^T + R \end{bmatrix}\\
\mathbf{X}_{ss} &= \mathbf{A}^d\mathbf{P}_{ss}^{L}(\mathbf{A}^{d+1})^T + \mathbf{L}_{ss} - \mathbf{A}^d\mathbf{P}_{ss}^{L}(\mathbf{A}^d)^T\\
\mathbf{L}_{ss} &= \mathbf{A}\mathbf{L}_{ss}\mathbf{A}^T + \mathbf{A}\mathbf{Q}
\end{aligned}
$$

c) End of interval exceedance/fade event: This event is similar to the end of interval up/downcrossing event, only differing in the fact that an exceedance is a one-dimensional event. Therefore, the level crossing condition reduces to Eqn. 41.

$$C_{exact} \triangleq \{|x_{k+d}| > L\} \qquad (41)$$

The probability $P(C_{exact})$ is shown in Eqn. 42, and is also equivalent to the p-value. This important relationship has practical value and will be discussed later.

$$
\begin{aligned}
P(C_{exact}) &= P(|x_{k+d}| > L) \qquad (42)\\
&= 2\Phi\left(-\frac{L}{\sqrt{\mathbf{C}\mathbf{P}_{ss}^{L}\mathbf{C}^T + R}}\right)\\
&= p
\end{aligned}
$$

d) At least one exceedance/fade event within an interval: This event was introduced by Kerr [10] as a problem for study, and can be represented as shown in Eqn. 43.

$$C_{exact} \triangleq \{|x_k| > L\} \bigcup \left[\bigcup_{j=1}^{d}\left[\bigcap_{i=0}^{j-1}|x_{k+i}| < L, |x_{k+j}| > L\right]\right] \qquad (43)$$

All exceedance sub-events in the expression are mutually exclusive. The expression represents all combinations of exceedances within the given prediction window, d. The probability $P(C_{exact})$ is shown in Eqn. 44.

$$
\begin{aligned}
P(C_{exact}) &= P(|x_k| > L) + \ldots\\
&\quad \sum_{j=1}^{d} P(\bigcap_{i=0}^{j-1}|x_{k+i}| < L, |x_{k+j}| > L) \qquad (44)\\
&= p + \sum_{j=1}^{d} P(\bigcap_{i=0}^{j-1}|x_{k+i}| < L, |x_{k+j}| > L)
\end{aligned}
$$

$$
\begin{aligned}
&= p + \sum_{j=1}^{d}\int_{L}^{\infty}\int_{-L}^{L}\cdots\int_{-L}^{L}\mathcal{N}(\mathbf{x}_j;\mu_{\mathbf{x}_j},\Sigma_{\mathbf{x}_j})\ d\mathbf{x}_j\\
&\quad + \sum_{j=1}^{d}\underbrace{\int_{-\infty}^{-L}\int_{-L}^{L}\cdots\int_{-L}^{L}}_{j+1}\mathcal{N}(\mathbf{x}_j;\mu_{\mathbf{x}_j},\Sigma_{\mathbf{x}_j})\ d\mathbf{x}_j
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{x}_j &= \begin{bmatrix} x_k \\ \vdots \\ x_{k+j} \end{bmatrix}, \mu_{\mathbf{x}_j} = \begin{bmatrix} \mathbf{C}\mu_{\mathbf{q}} \\ \vdots \\ \mathbf{C}\mu_{\mathbf{q}} \end{bmatrix} \in \mathbb{R}^{j+1}\\
\Sigma_{\mathbf{x}_j}(i_1,i_2) &= \begin{cases} \mathbf{C}\mathbf{P}_{ss}^{L}\mathbf{C}^T + \mathbf{R} & 1 \le i_1 = i_2 \le j+1 \\ \mathbf{C}\mathbf{X}_{ss}\mathbf{C}^T & 1 \le i_1 \ne i_2 \le j+1 \end{cases}\\
\mathbf{X}_{ss} &= \mathbf{A}^{i_1}\mathbf{P}_{ss}^{L}(\mathbf{A}^{i_2})^T + \mathbf{L}_{ss} - \mathbf{A}^{i_1}\mathbf{L}_{ss}(\mathbf{A}^{i_1})^T\\
\mathbf{L}_{ss} &= \mathbf{A}\mathbf{L}_{ss}\mathbf{A}^T + \mathbf{A}^{i_2-i_1}\mathbf{Q} \text{ where } i_1 < i_2\\
\mu_{\mathbf{q}} &\triangleq E[\mathbf{q}_k]
\end{aligned}
$$

e) At least one up/downcrossing event within an interval: This event is identical to the previous level-crossing event except that two-dimensional up/downcrossings are considered in lieu of exceedances. The expression describing this event is actually simpler, as in Eqn. 45.

$$C_{exact} \triangleq \bigcup_{j=1}^{d}\left\{\bigcap_{i=0}^{j-1}|x_{k+i}| < L, |x_{k+j}| > L\right\} \qquad (45)$$

The probability $P(C_{exact})$ is shown in Eqn. 46, with identical definitions of $\mathbf{x}_j$, $\mu_{\mathbf{x}_j}$, $\Sigma_{\mathbf{x}_j}$ as in the previous case.

$$
\begin{aligned}
P(C_{exact}) &= \sum_{j=1}^{d} P(\bigcap_{i=0}^{j-1}|x_{k+i}| < L, |x_{k+j}| > L) \qquad (46)\\
&= \sum_{j=1}^{d}\int_{L}^{\infty}\int_{-L}^{L}\cdots\int_{-L}^{L}\mathcal{N}(\mathbf{x}_j;\mu_{\mathbf{x}_j},\Sigma_{\mathbf{x}_j})\ d\mathbf{x}_j\\
&\quad + \sum_{j=1}^{d}\underbrace{\int_{-\infty}^{-L}\int_{-L}^{L}\cdots\int_{-L}^{L}}_{j+1}\mathcal{N}(\mathbf{x}_j;\mu_{\mathbf{x}_j},\Sigma_{\mathbf{x}_j})\ d\mathbf{x}_j
\end{aligned}
$$

For each of the five listed cases: *a-e*, we will study and compare the results of three distinct types of alarm systems, previously introduced. The alarm system metrics of interest for this application are the ROC curve statistics: the true and false positive rates. Eqns. 47-48 summarize the formulae necessary to compute these statistics. Notice that the true positive rate shown in Eqn. 47 is identical to Eqn. 29, for the probability of correct alarm, also know as recall. There is overlap here, and there is also usefulness in looking at different alarm system metrics, which is discussed at length in [16]. In general, as long as the following three probability computations are performed: $P(C_{exact})$, $P(A_{approx})$, and $P(C_{exact}, A_{approx})$, any relevant alarm system metric can be derived.

True positive rate:

$$P(C_{exact}|A_{approx}) = \frac{P(C_{exact}, A_{approx})}{P(A_{approx})} \qquad (47)$$

False positive rate:

$$P(A_{approx}|C'_{exact}) = \frac{P(C'_{exact}, A_{approx})}{P(C'_{exact})} \qquad (48)$$

a) "Redlines" or Simple Alarm System: In order to introduce the use of redlines we first make the distinction between the critical level, L, and the redline, denoted as L_A. The critical level represents the threshold above which damage or some significant decrease in quality of a behavior or process may potentially occur. There are some cases in which this critical level is not known, have not been designed a priori, or when known critical levels yield alarm systems that are practically infeasible. The latter case usually occurs when the thresholds are set to levels so extreme that the resulting probability computations default to null values. As such, sometimes it is beneficial to use values that are based upon statistical outlier detection and hypothesis testing via the p-value. The relationship between the critical level, L, and the p-value is shown in Eqn. 49

$$L_{equiv} = \Phi^{-1}\left(1 - \frac{p}{2}\right)\sqrt{\mathbf{C}\mathbf{P}_{ss}^{L}\mathbf{C}^T + R} \tag{49}$$

The redline value given by L_A is a different parameter than the critical level, L, and essentially acts as a design parameter with which to tune the alarm system sensitivity. Its value is the level at which an alarm would literally sound, and whose selection may be performed manually via brute force gridding, or related to the use of a log-likelihood based method. Using the log-likelihood based method provides an alternative method of design for the alarm system that is essentially equivalent to choosing a redline value via the χ_1^2 distribution using the following equations:

$$P(\log(p(x_k; 0, \mathbf{C}\mathbf{P}_{ss}^{L}\mathbf{C}^T + R)) < l) =$$
$$1 - \chi_1^2\left(-2\left[l + \frac{1}{2}\log(2\pi) + \log\sqrt{\mathbf{C}\mathbf{P}_{ss}^{L}\mathbf{C}^T + R}\right]\right) =$$
$$2\Phi\left(-\sqrt{-2\left[l + \frac{1}{2}\log(2\pi) + \log\sqrt{\mathbf{C}\mathbf{P}_{ss}^{L}\mathbf{C}^T + R}\right]}\right) =$$
$$2\Phi\left(\frac{-L_{A_{equiv}}}{\sqrt{\mathbf{C}\mathbf{P}_{ss}^{L}\mathbf{C}^T + R}}\right) = 2\Phi\left(\frac{-L_{A_{equiv}}}{\sigma}\right)$$

Therefore, Eqn. 50 represents the equivalent value for L_A, where l is a design log-likelihood based threshold.

$$L_{A_{equiv}} = \sqrt{-2\sigma^2\left[l + \frac{1}{2}\log(2\pi) + \log\sigma\right]} \tag{50}$$

In general, using the log-likelihood value as the basis for outlier detection allows for greater accessibility of infinitesimally small values of the significance level (i.e., $p = 1\times10^{-8}$) for the corresponding hypothesis-based decision test. A more thorough discussion on the use of this approach for alarm systems can be found in [9], [19] and [20]. Note that an equivalent p-value for the design parameter, L_A, can also be found with Eqn. 49. The probability of alarm for a redline alarm system regardless of level-crossing type is given by Eqn. 51. Note that the alarm system never uses any predicted future values, only the value at the current time, such that $A_{approx} \triangleq \{|x_k| > L_A\}$.

$$\begin{aligned} P(A_{approx}) &= P(|x_k| > L_A) \\ &= 2\Phi\left(-\frac{L_A}{\sqrt{\mathbf{C}\mathbf{P}_{ss}^{L}\mathbf{C}^T + R}}\right) \end{aligned} \tag{51}$$

As mentioned previously, the probability of the critical event, $P(C_{exact})$, as well as the probability of alarm, $P(A_{approx})$, and the joint probability of the two, $P(C_{exact}, A_{approx})$, suffice to compute any relevant alarm system metrics. Since we already have the necessary formulae for computing $P(C_{exact})$ and $P(A_{approx})$, we provide the detailed formulae for $P(C_{exact}, A_{approx})$ in Appendix I.

b) Predictive Alarm System: The predictive alarm system uses a similar fixed, static threshold, L_A, akin to the redline method. However, rather than the current process value, x_k, being used, predicted future values, $\hat{x}_{k+d|k}$, are compared to the alarm level, L_A. The probability of alarm for a predictive alarm system regardless of level-crossing type is given by Eqn. 52. For the case of the end of interval up/downcrossing event, the predicted future value of $\hat{x}_{k+d+1|k}$ is used in lieu of $\hat{x}_{k+d|k}$.

$$\begin{aligned} P(A_{approx}) &= P(|\hat{x}_{k+d|k}| > L_A) \\ &= 2\Phi\left(-\frac{L_A}{\sqrt{\mathbf{C}\mathbf{A}^d(\mathbf{P}_{ss}^{L} - \hat{\mathbf{P}}_{ss}^{R})(\mathbf{A}^d)^T\mathbf{C}^T}}\right) \end{aligned} \tag{52}$$

We present the formulae for computing the probability of $P(C_{exact}, A_{approx})$ for all level-crossing events of interest in Appendix II.

c) Optimal Alarm System: The optimal alarm system uses a concept introduced earlier, by defining the alarm region, A_{exact}, as follows:

$$A_{exact} \triangleq \{\hat{\mathbb{X}} : P(C_{exact}|\{x_0, \ldots, x_k\}) \geq P_b\}$$

where $\hat{\mathbb{X}}$ is a vector of all predicted future process values, i.e., $\hat{x}_{k|k}$, $\hat{x}_{k+d|k}$ that correspond to the future time steps in the definition of the critical event, C_{exact}. In this case, there is no fixed, static threshold, L_A, akin to the redline or predictive methods to act as a design parameter. However, the border probability, P_b, acts in place of L_A as an alarm system design parameter. As seen previously for the thermal sensation complaint application, the alarm region can be approximated easily by using a variety of methods which use inequalities involving predicted future process values, i.e., $\hat{x}_{k+d|k}$. Depending on the type of event under consideration, the approximation will vary. However, in general there are two types of approximations that can be made for the spacecraft propulsion system anomaly detection application.

Unlike the probability of alarm for a redline or predictive alarm system, the optimal alarm system's probability of alarm is dependent on level-crossing type. As such, detailed formulae for $P(A_{approx})$ and $P(C_{exact}, A_{approx})$ for all level-crossing events of interest are provided in Appendix III. However, as a precursor, we must derive the two types of approximations

to the alarm regions required for computation of these probabilities. We begin with an up/downcrossing event spanning an interval, whose alarm condition is shown in Eqn. 53.

$$P(|x_k| < L, |x_{k+d}| > L | x_0, \ldots, x_k) \geq P_b \quad (53)$$

$$\lim_{|\hat{x}_{k+d|k}| \to \infty} P(|x_k| < L, |x_{k+d}| > L | x_0, \ldots, x_k) = P(|x_k| < L | x_0, \ldots, x_k)$$

$$\Updownarrow$$

$$P(|x_k| < L | x_0, \ldots, x_k) \geq P_b \quad (54)$$

The first approximation can be thought of as a "closed-form" approximation, meaning that at least one of the resulting inequalities involving $\hat{x}_{k|k}$ and $\hat{x}_{k+d|k}$ can be expressed directly as a function of the model parameters. The second approximation can be found as a one-dimensional scalar nonlinear root-finding problem, [21]. The asymptote corresponding to $\hat{x}_{k|k}$ for the first approximation can be found in closed form by taking the intersection of the inequalities in Eqns. 55-56, culminating in the alarm region represented by Eqn. 57. An additional approximation is introduced by taking the intersection of these two inequalities. These inequalities come from Eqn. 54, which is derived by finding the limiting distribution as the remaining dimension is marginalized by taking $\lim_{|\hat{x}_{k+d|k}| \to \infty}$.

$$P(x_k < L | x_0, \ldots, x_k) \geq P_b \quad (55)$$

$$P(x_k > -L | x_0, \ldots, x_k) \geq P_b \quad (56)$$

$$\Updownarrow$$

$$|\hat{x}_{k|k}| \leq L - \sqrt{V_{k|k}}\Phi^{-1}(P_b) = L_A^- \quad (57)$$

The closed-form approximations shown in Eqns. 53 and 57 are similar to the optimal two-dimensional alarm region approximations from the previous application found in Eqns. 22 and 23, respectively. The sole difference is that $|x_k|$ is used in place of x_k for the current example. The root-finding approximation can be found by solving for the zeros of $f(\hat{x}_{k|k})$ shown in Eqn. 58, given a particular value of P_b, yielding the asymptote for the alarm region corresponding to $\hat{x}_{k|k}$ such that $|\hat{x}_{k|k}| \leq L_A^-$ for $f(\hat{x}_{k|k}) \geq 0$. Therefore, no additional approximations are introduced beyond using the asymptotes themselves.

$$f(\hat{x}_{k|k}) = \Phi\left(\frac{L - \hat{x}_{k|k}}{\sqrt{V_{k|k}}}\right) - \Phi\left(\frac{-L - \hat{x}_{k|k}}{\sqrt{V_{k|k}}}\right) - P_b \quad (58)$$

When deriving the asymptote for $\hat{x}_{k+d|k}$, there is no limiting approximation as in Eqn. 54. As such, in order to make a closed-form approximation, we use an intuitive approach by taking the union of the following inequalities, culminating in the alarm region represented by Eqn. 59, again introducing an additional approximation.

$$P(x_{k+d} > L | x_0, \ldots, x_k) \geq P_b$$
$$P(x_{k+d} < -L | x_0, \ldots, x_k) \geq P_b$$

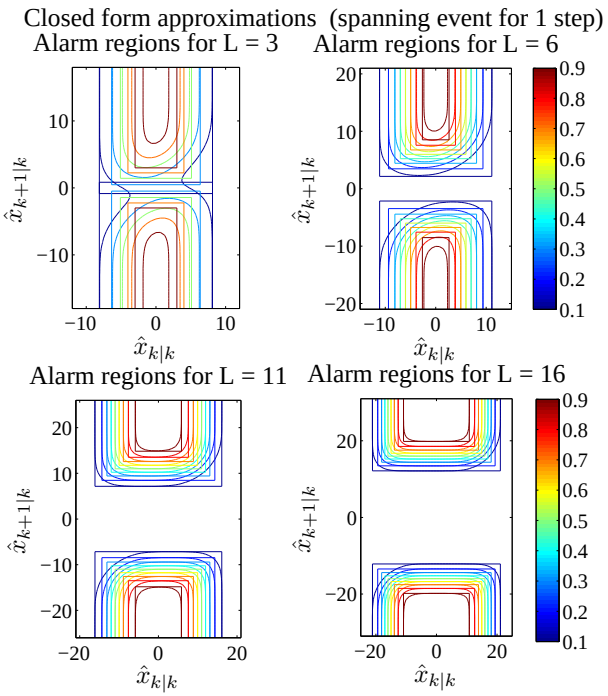

Fig. 3. Closed-form approximations for sample alarm regions

$$\Updownarrow$$

$$|\hat{x}_{k+d|k}| \geq L + \sqrt{V_{k+d|k}}\Phi^{-1}(P_b) = L_A^+ \quad (59)$$

Although this provides a closed-form approximation, there is inherently no mathematically defensible argument for doing so. As such, we provide the more mathematically correct approach, which is akin to the root-finding approximation method, by solving for the zeros of $f(\hat{x}_{k|k}, \hat{x}_{k+d|k})\ |_{\hat{x}_{k|k}=0}$ shown in Eqn. 60, given a particular value of P_b, yielding the asymptote for the alarm region corresponding to $\hat{x}_{k+d|k}$ such that $|\hat{x}_{k+d|k}| \geq L_A^+$ for $f(\hat{x}_{k|k}, \hat{x}_{k+d|k})\ |_{\hat{x}_{k|k}=0} \geq 0$.

$$f(\hat{x}_{k|k}, \hat{x}_{k+d|k}) = P(|x_k| < L, |x_{k+d}| > L | x_0, \ldots, x_k) - P_b \quad (60)$$

$$= \left(\int_{-L}^{L}\int_{L}^{\infty} + \int_{-L}^{L}\int_{-\infty}^{-L}\right) \mathcal{N}(\mathbf{x}; \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})\ d\mathbf{x} - P_b$$

where

$$\mathbf{x} = \begin{bmatrix} x_{k|k} \\ x_{k+d|k} \end{bmatrix}, \mu_{\mathbf{x}} = \begin{bmatrix} \hat{x}_{k|k} \\ \hat{x}_{k+d|k} \end{bmatrix} \in \mathbb{R}^2$$

and $\Sigma_{\mathbf{x}}(i_1, i_2) =$

$$\begin{cases} \mathbf{C}\hat{\mathbf{P}}_{ss}^R\mathbf{C}^T + R & i_1 = i_2 = 1 \\ \mathbf{C}\left[\mathbf{A}^d(\hat{\mathbf{P}}_{ss}^R - \mathbf{P}_{ss}^L)\left(\mathbf{A}^d\right)^T + \mathbf{P}_{ss}^L\right]\mathbf{C}^T + R & i_1 = i_2 = 2 \\ \mathbf{C}\left[\hat{\mathbf{P}}_{ss}^R - \left(\mathbf{A}^T\right)^d\right]\mathbf{C}^T & i_1 \neq i_2 \end{cases}$$

Figs. 3 and 4 depict the qualitative nature of both the closed-form and root-finding approximations to generic sample alarm

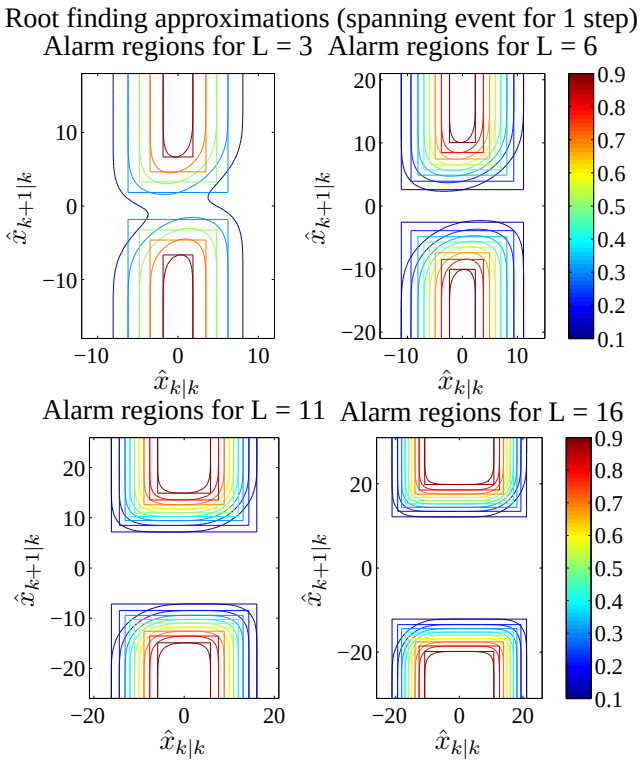

Fig. 4. Root-finding approximations for sample alarm regions

regions for values of P_b shown in gradations of 0.1, respectively. These sample alarm regions are based on various levels as shown in the titles of the subplots in the figures and the following covariance matrix:

$$\Sigma_{\mathbf{x}} = \begin{bmatrix} 16 & 6 \\ 6 & 9 \end{bmatrix}$$

This simple covariance matrix was generated by forming a covariance matrix based upon the following standard deviations of the random vector $\mathbf{x}$:

$$\begin{bmatrix} 4 \\ 3 \end{bmatrix}$$

The correlation coefficient used is $\rho = 0.5$ in order to simulate the eccentricity that would most likely be introduced in real data so as to illustrate that the data are not independent, but correlated.[3]

These alarm regions are based upon Eqn. 53, and the closed-form and root finding approximations are shown in Figs. 3 and 4, respectively. Looking very closely at the figures, one can see the exact plots are represented by the curved regions, and are approximated by the rectangular bounds that represent both the closed-form and root finding approximations. Improvement in the approximations can be discerned by closely examining the tightness of the rectangular bounds. Notice that the approximations improve as L increases in both Figs. 3 and 4 due to the eccentricity of the alarm regions. In contrast, for smaller values of L, and as P_b increases, the closed-form approximation worsens as seen in Fig. 3. The same is true for the root-finding method, however, as seen in Fig. 4, the asymptotic bounds are tighter. This implies that the alarm region represented by Eqns. 57 and 59 (closed-form approximation) provide a poorer approximation to the alarm region than with Eqns. 58 and 60, when using the root-finding method.

[3]As such, these sample alarm regions are again generated solely for illustrative purposes and are not based on models generated from experimental data.

The alarm condition for an end of interval up/downcrossing event is shown in Eqn. 61.

$$P(|x_{k+d}| < L, |x_{k+d+1}| > L | x_0, \ldots, x_k) \geq P_b \tag{61}$$

We will forgo the derivations of the two types of approximations to the alarm regions, as they are similar to those for the up/downcrossing event spanning an interval. As such, the resulting closed form alarm regions are shown in Eqns. 62 and 63.

$$|\hat{x}_{k+d|k}| \leq L - \sqrt{V_{k+d|k}}\Phi^{-1}(P_b) = L_A^- \tag{62}$$

$$|\hat{x}_{k+d+1|k}| \geq L + \sqrt{V_{k+d+1|k}}\Phi^{-1}(P_b) = L_A^+ \tag{63}$$

The resulting root-finding alarm regions are governed by the same inequalities in Eqns. 62 and 63. L_A^- and L_A^+ can be found by solving for $f(\hat{x}_{k+d|k}) \geq 0$ and $f(\hat{x}_{k+d|k}, \hat{x}_{k+d+1|k})\ |_{\hat{x}_{k+d|k}=0} \geq 0$, where

$$f(\hat{x}_{k+d|k}) = \Phi\left(\frac{L - \hat{x}_{k+d|k}}{\sqrt{V_{k+d|k}}}\right) - \Phi\left(\frac{-L - \hat{x}_{k+d|k}}{\sqrt{V_{k+d|k}}}\right) - P_b$$

and $f(\hat{x}_{k+d|k}, \hat{x}_{k+d+1|k}) =$

$$\begin{aligned} & P(|x_{k+d}| < L, |x_{k+d+1}| > L | x_0, \ldots, x_k) - P_b \\ = & \left(\int_{-L}^{L}\int_{L}^{\infty} + \int_{-L}^{L}\int_{-\infty}^{-L}\right) \mathcal{N}(\mathbf{x}; \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})\ d\mathbf{x} - P_b \end{aligned}$$

where

$$\mathbf{x} = \begin{bmatrix} x_{k+d|k} \\ x_{k+d+1|k} \end{bmatrix}, \mu_{\mathbf{x}} = \begin{bmatrix} \hat{x}_{k+d|k} \\ \hat{x}_{k+d+1|k} \end{bmatrix} \in \mathbb{R}^2$$

and $\Sigma_{\mathbf{x}}(i_1, i_2) =$

$$\begin{cases} \mathbf{C}\mathbf{P}_{k+d|k}\mathbf{C}^T + R & i_1 = i_2 = 1 \\ \mathbf{C}\mathbf{P}_{k+d+1|k}\mathbf{C}^T + R & i_1 = i_2 = 2 \\ \mathbf{C}\mathbf{X}_{ss}\mathbf{C}^T & i_1 \neq i_2 \end{cases}$$

where

$$\begin{aligned} \mathbf{P}_{k+d|k} &\triangleq \mathbf{A}^d(\hat{\mathbf{P}}_{ss}^R - \mathbf{P}_{ss}^L)\left(\mathbf{A}^d\right)^T + \mathbf{P}_{ss}^L \\ \mathbf{P}_{k+d+1|k} &\triangleq \mathbf{A}^{d+1}(\hat{\mathbf{P}}_{ss}^R - \mathbf{P}_{ss}^L)\left(\mathbf{A}^{d+1}\right)^T + \mathbf{P}_{ss}^L \\ \mathbf{X}_{ss} &= \mathbf{A}^d\left[\hat{\mathbf{P}}_{ss}^R(\mathbf{A})^T - \mathbf{L}_{ss}\right]\left(\mathbf{A}^d\right)^T + \mathbf{L}_{ss} \end{aligned}$$

and $\mathbf{L}_{ss} = \mathbf{A}\mathbf{L}_{ss}\mathbf{A}^T + \mathbf{A}\mathbf{Q}$

The alarm condition for an end of interval exceedance/fade event is shown in Eqn. 64.

$$P(|x_{k+d}| > L | x_0, \ldots, x_k) \geq P_b \tag{64}$$

Again, we will forgo the derivations of the two types of approximations to the alarm regions, as they are similar to those for the previous cases. As such, the resulting closed form alarm region is shown in Eqn. 65.

$$|\hat{x}_{k+d|k}| \geq L + \sqrt{V_{k+d|k}}\Phi^{-1}(P_b) = L_A^+ \tag{65}$$

The resulting root-finding alarm regions are governed by the same inequality in Eqn. 65. L_A^+ can be found by solving for $f(\hat{x}_{k+d|k}) \geq 0$, where

$$f(\hat{x}_{k+d|k}) = \Phi\left(\frac{\hat{x}_{k+d|k} - L}{\sqrt{V_{k+d|k}}}\right) + \Phi\left(\frac{-\hat{x}_{k+d|k} - L}{\sqrt{V_{k+d|k}}}\right) - P_b$$

The alarm condition for at least one exceedance/fade event within an interval is $P(C_{exact}|\mathcal{D}) = P(C_{exact}|x_0, \ldots, x_k) \geq P_b$, and is expanded in Eqn. 66.

$$P(|x_k| > L|\mathcal{D}) + \sum_{j=1}^{d} P(\bigcap_{i=0}^{j-1} |x_{k+i}| < L, |x_{k+j}| > L|\mathcal{D}) \geq P_b \tag{66}$$

To determine the approximations to the alarm regions, we can again use either the closed-form or root-finding methods. Using the closed-form approximation, we intuitively use the same logic as in previous cases. However, we have the added caveat of determining the alarm region corresponding to the sum of several conditional probabilities on the left hand side of the inequality of Eqn. 66, unlike before. As such, we shall approximate this inequality by the union of $d+1$ distinct inequalities. These inequalities will be formed by conditional probabilities of subevents on the left hand side of the inequalities, and the same design parameter, P_b, on the right hand side for each inequality. It is reasonable to make this assumption because all subevents are disjoint, and therefore can be considered independently. The same mathematical or intuitive arguments as used for Eqns. 53 - 59 apply to each of these inequalities as well. As such, the closed form approximation can be represented for all asymptotes corresponding to $|\hat{x}_{k+i|k}|, \forall i \geq 0$, as shown in Eqns. 67 and 68.

$$A_{approx} = \bigcup_{i=0}^{d} |\hat{x}_{k+i|k}| \geq \underbrace{L + \sqrt{V_{k+i|k}}\Phi^{-1}(P_b)}_{L_{A_i}^+} \tag{67}$$

Using the root-finding approximation, the same union of inequalities applies, as in Eqn. 68.

$$A_{approx} = \bigcup_{i=0}^{d} |\hat{x}_{k+i|k}| \geq L_{A_i}^+ \tag{68}$$

However, the alarm region can now be found by the inequality $f(\hat{\mathbb{X}}) \mid_{\hat{\mathbb{X}} \setminus \hat{x}_{k+i|k}=0} \geq 0$, where the asymptotes, $L_{A_i}^+$ can be found by solving for the zeros of $f(\hat{\mathbb{X}}) = P(C_{exact}|\mathcal{D}) - P_b =$

$$P(|x_k| > L|\mathcal{D}) + \sum_{j=1}^{d} P(\bigcap_{i=0}^{j-1} |x_{k+i}| < L, |x_{k+j}| > L|\mathcal{D}) - P_b$$

$$= \Phi\left(\frac{\hat{x}_{k|k} - L}{\sqrt{V_{k|k}}}\right) + \Phi\left(\frac{-\hat{x}_{k|k} - L}{\sqrt{V_{k|k}}}\right) + \ldots$$

$$\sum_{j=1}^{d} \int_{L}^{\infty} \int_{-L}^{L} \ldots \int_{-L}^{L} \mathcal{N}(\mathbf{x}_j; \mu_{\mathbf{x}_j}, \Sigma_{\mathbf{x}_j}) d\mathbf{x}_j + \ldots$$

$$\sum_{j=1}^{d} \underbrace{\int_{-\infty}^{-L} \int_{-L}^{L} \ldots \int_{-L}^{L}}_{j+1} \mathcal{N}(\mathbf{x}_j; \mu_{\mathbf{x}_j}, \Sigma_{\mathbf{x}_j})\ d\mathbf{x}_j$$

$$\begin{aligned}
\mathbf{x}_j &= \begin{bmatrix} x_{k|k} \\ \vdots \\ x_{k+j|k} \end{bmatrix}, \mu_{\mathbf{x}_j} = \begin{bmatrix} \hat{x}_{k|k} \\ \vdots \\ \hat{x}_{k+j|k} \end{bmatrix} \in \mathbb{R}^{j+1} \\
\Sigma_{\mathbf{x}}(i_1, i_2) &= \begin{cases} \mathbf{C}\mathbf{P}_{k+i_1|k}\mathbf{C}^T + R & i_1 = i_2 \\ \mathbf{C}\mathbf{X}_{ss}\mathbf{C}^T & i_1 \neq i_2 \end{cases}
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{P}_{k+i_1|k} &\triangleq \mathbf{A}^{i_1}(\hat{\mathbf{P}}_{ss}^R - \mathbf{P}_{ss}^L)\left(\mathbf{A}^{i_1}\right)^T + \mathbf{P}_{ss}^L \\
\mathbf{X}_{ss} &= \mathbf{A}^{i_1}\left[\hat{\mathbf{P}}_{ss}^R\left(\mathbf{A}^{i_2-i_1}\right)^T - \mathbf{L}_{ss}\right]\left(\mathbf{A}^{i_1}\right)^T + \mathbf{L}_{ss}
\end{aligned}$$

and $\mathbf{L}_{ss} = \mathbf{A}\mathbf{L}_{ss}\mathbf{A}^T + \mathbf{A}\mathbf{Q}$

This alarm region may span many dimensions, but can also be shown in two or three dimensions. As such, we provide Figs. 5 and 6, which illustrate the qualitative nature of both the closed-form and root-finding approximations in two dimensions for sample alarm regions, as in Figs. 3 and 4. Again, we see that the root-finding approximation provides a better bound on the alarm region than the closed-form approximation. We also show a sample three dimensional alarm region in Fig. 7, for illustrative purposes. The same approximations can be used to form a hypercube, outside of which integrations to compute relevant alarm statistics may be performed.

Finally, for at least one up/downcrossing event within an interval, the alarm condition, $P(C_{exact}|\mathcal{D}) = P(C_{exact}|x_0, \ldots, x_k) \geq P_b$, is expanded in Eqn. 69.

$$\sum_{j=1}^{d} P(\bigcap_{i=0}^{j-1} |x_{k+i}| < L, |x_{k+j}| > L|\mathcal{D}) \geq P_b \tag{69}$$

To determine the approximations to the alarm regions, we can again use either the closed-form or root-finding methods. Both approximations can be represented identically for all asymptotes corresponding to $|\hat{x}_{k+i|k}|, \forall i \geq 1$, as shown previously in Eqns. 67 and 68. However, here these approximations are good only for the asymptotes $\forall i \geq 1$ as distinct from $\forall i \geq 0$ with at least one exceedance/fade event within an interval. The alarm region approximations for this subspace are represented by Eqns. 70 and 71, for the closed-form and root-finding methods, respectively.

$$\bigcup_{i=1}^{d} |\hat{x}_{k+i|k}| \geq \underbrace{L + \sqrt{V_{k+i|k}}\Phi^{-1}(P_b)}_{L_{A_i}^+} \tag{70}$$

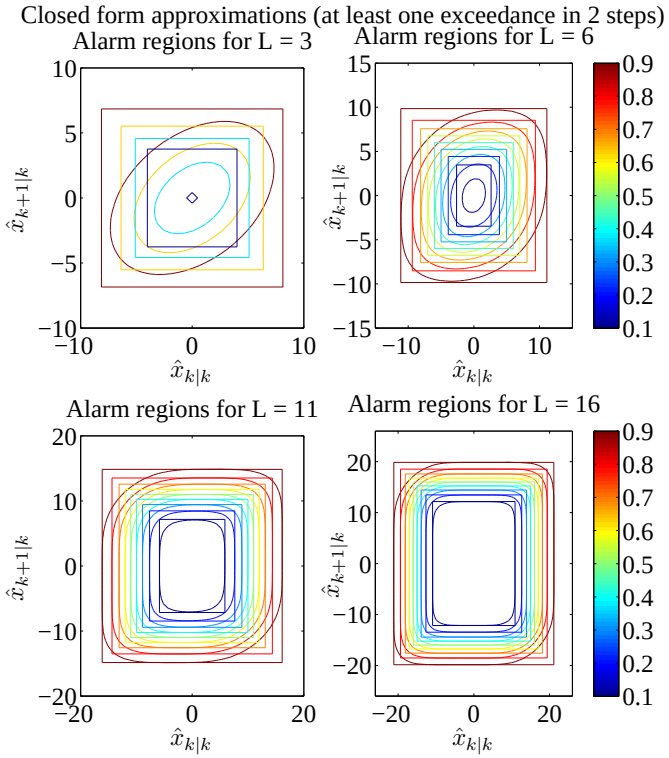

Fig. 5. Closed-form approximations for multiple exceedance alarm region

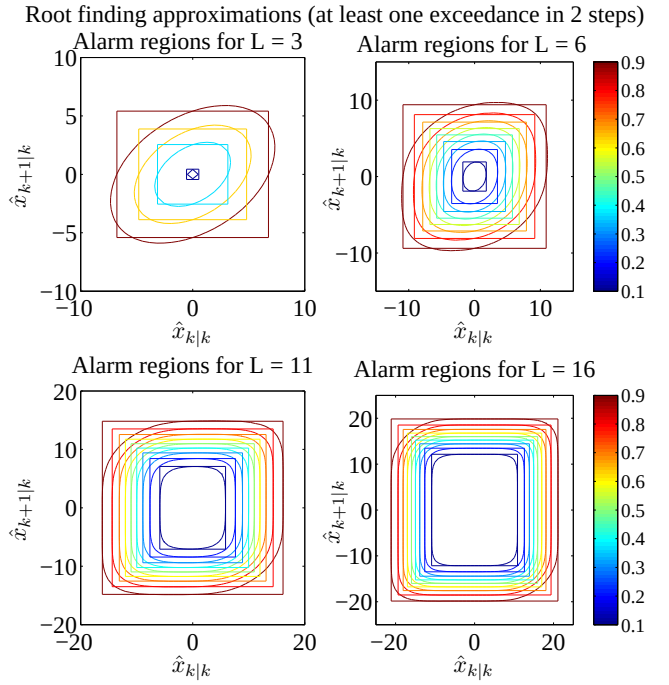

Fig. 6. Root-finding approximations for multiple exceedance alarm region

$$\bigcup_{i=1}^{d} |\hat{x}_{k+i|k}| \geq L_{A_i}^{+} \tag{71}$$

For the remaining asymptote corresponding to $|\hat{x}_{k|k}|$, we provide the same two alternatives again for either the closed-form or root-finding approximations based upon the derivation provided below.

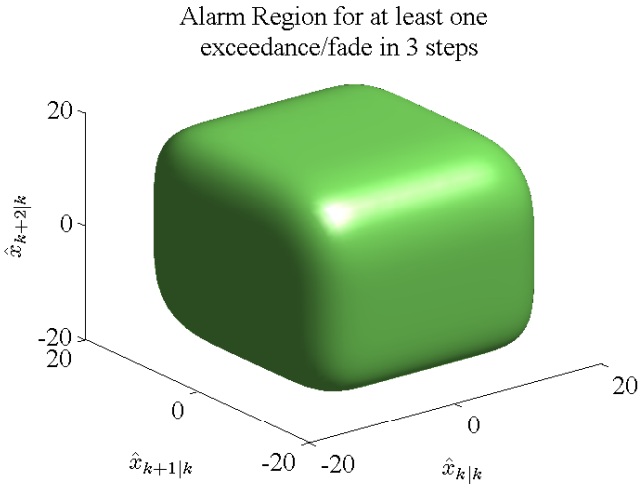

Fig. 7. Multiple exceedance alarm region in 3 dimensions

$$\lim_{\hat{\mathbb{X}}\setminus|\hat{x}_{k|k}|\to\infty} \sum_{j=1}^{d} P(\bigcap_{i=0}^{j-1} |x_{k+i}| < L, |x_{k+j}| > L|\mathcal{D}) =$$

$$P(|x_k| < L|x_0, \ldots, x_k)$$

$$\Updownarrow$$

$$P(|x_k| < L|x_0, \ldots, x_k) \geq P_b$$

This follows from the same logic that leads up to Eqn. 35 for the previously addressed thermal sensation complaint application. As such, we may use Eqns. 55-57 to define the closed-form or root-finding approximations and the asymptote for $|\hat{x}_{k|k}|$, previously used for the derivation of Eqn. 57 in the beginning of this section. For the closed form approximation, it is given by the same inequality, shown again in Eqn. 72.

$$|\hat{x}_{k|k}| \leq L - \sqrt{V_{k|k}}\Phi^{-1}(P_b) = L_A^{-} \tag{72}$$

Furthermore, the root-finding approximation can again be found by solving for the zeros of $f(\hat{x}_{k|k})$ shown below, just as in Eqn. 58 provided previously. Given a particular value of P_b, the asymptote for the alarm region corresponding to $\hat{x}_{k|k}$ is defined by L_A^{-}, where $|\hat{x}_{k|k}| \leq L_A^{-}$ for $f(\hat{x}_{k|k}) \geq 0$.

$$f(\hat{x}_{k|k}) = \Phi\left(\frac{L - \hat{x}_{k|k}}{\sqrt{V_{k|k}}}\right) - \Phi\left(\frac{-L - \hat{x}_{k|k}}{\sqrt{V_{k|k}}}\right) - P_b$$

The combined alarm region is the intersection of Eqns. 71 and 72, yielding Eqn. 73.

$$A_{approx} = \bigcup_{i=1}^{d} \left[|\hat{x}_{k+i|k}| \geq L_{A_i}^{+}\right] \bigcap \left[|\hat{x}_{k|k}| \leq L_A^{-}\right] \tag{73}$$

This alarm region is the last of the optimal regions to be discussed. However, there may be many more that can be

defined from within the class of level-crossing events having a fixed threshold, using the same techniques. This particular event, in which at least one up/downcrossing occurs within an interval, may also span many dimensions as in the previous example. We show a sample three dimensional alarm region in Fig. 8, for illustrative purposes. The approximations provided for this alarm region can be used to form a rectangularized bound, again outside of which integrations to compute relevant alarm statistics may be performed. In two dimensions, this event is identical to the first one covered in this section, an up/downcrossing event spanning an interval. The corresponding relevant approximations are shown in Figs. 3 and 4.

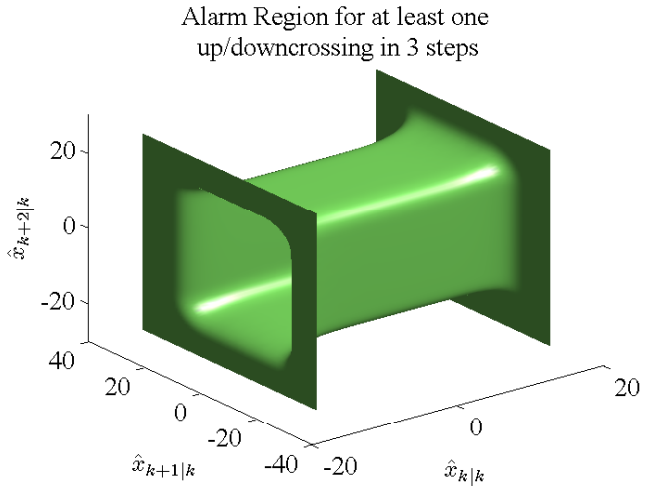

Fig. 8. Multiple up/downcrossing alarm region in 3 dimensions

V. Examples

For both of the examples to be used as a demonstration testbed for the theory presented in the previous section, we use variants of data-driven methods to arrive at statistically viable models. It is possible and in some cases preferable to use a model that incorporates physics and is based upon first principles. In these situations, model fidelity is of paramount importance, so that simulation of the model results in a realization that both quantitatively and qualitatively resembles real system behavior. These models are often of use when implementing algorithms that automatically take corrective action, or respond appropriately to caution and warning alarm signals. In some cases, philosophical or political drivers mandate the use of such models, when intuitive and physical explanations are required for the implementation of alarm systems.

However, developing such models often requires extensive labor and expert domain knowledge. Invoking the data-driven approach ameliorates this requirement. Furthermore, data-driven models are sufficient for our purposes here because statistical characterization is a reasonable first step in application of alarm systems based upon control system error. More extensive models may be required when developing automated response strategies as in [3]. As such, we present models based upon the data-driven approach at different stages for both the spacecraft propulsion anomaly detection and the thermal sensation complaint example.

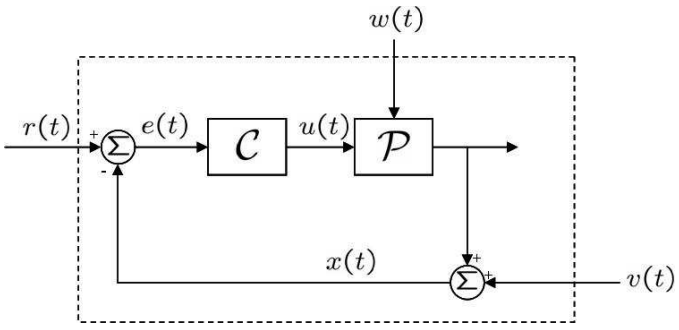

Fig. 9. Closed-Loop Control System Block Diagram

A. Spacecraft Propulsion System Anomaly Detection Application

The model used for the spacecraft propulsion system anomaly detection application will be briefly reviewed here. More extensive details can be found in [9]. This model is based upon control system error, with a fixed critical threshold as the primary indicator of criticality. The data used for training of the model is also discussed in detail in [9]. For the example presented here, the model is based upon the reduced dataset discussed in [9] that eliminates certain tests based upon functional categorization. There are several motivations behind training a linear dynamic system using control system error for the spacecraft propulsion system anomaly detection application. One relates to the fact that the data requirements are quite modest. The training data is univariate, i.e., $n = 1$, and represents the difference between the commanded throttle and actual throttle. In control systems terminology, this is the control system error, $e(t)$, traditionally used as the input to a controller, as shown in Fig. 9. At the same time, we can provide for a richer description of the dynamics of the data.

We also appeal to the use of control system error against limit checks or redlines in the design of various alarm systems. Disturbances that influence a control system during nominal operation may cause a threshold to be exceeded. However, other non-environmental disturbances may represent subtleties in the dynamics of the system being controlled. These anomalous excursions may potentially stem from latent faults in the controlled system that are precursors to incipient failures, and may eventually manifest themselves in a more serious fashion. As such, the control system error provides an excellent parameter for monitoring in the face of latent faults that may present themselves in a more nuanced manner. This measure may serve as a more advanced technique to complement algorithms that use direct sensor measurement which have more physically intuitive interpretations when applying the limit check paradigm.

There are two primary control systems that operate in support of the Space Shuttle Main Engine (SSME), the spacecraft propulsion system which is the driver for all of the models developed in this section. One is the throttle control system, which regulates the main combustion chamber

pressure. The other major control system that functionally supports the SSME is the mixture ratio control system. This system maintains the oxidizer/fuel mixture ratio at a desired level. We use the throttle control system error due to the commanded throttle qualitatively being the apparent driver for so many other sensor readings.

In Fig. 9, the closed-loop control system representation illustrates that the actual throttle level, $x(t)$, is subtracted from the desired or commanded throttle level, $r(t)$, to obtain control system error, $e(t) = r(t) - x(t)$. The block labelled $\mathcal{C}$ represents the controller, which we can nominally assume to be a very simple PI (proportional-integral) controller. The PI controller takes the control system error and computes the appropriate actuation to deliver to the plant, labelled as block $\mathcal{P}$. The plant may be subject to input noise, $w(t)$, which is introduced directly into the state dynamics. Finally, as the feedback loop is closed, measurement noise, $v(t)$, may be additively introduced to the output of the plant to form $x(t)$, used by the control system.

There are several transfer functions that can be formed from the closed-loop state dynamics. The one that we are most interested in from the machine learning standpoint is $TF_{w\rightarrow e}$, or the closed loop dynamics that represent the transfer function from input noise to error. Because the data available to us for training is the control system error, $e(t) = r(t) - x(t)$, we can reformulate the dynamics of the closed-loop feedback control system into a standard representation that can be treated as an unsupervised problem in machine learning (i.e., using output observations only). This is performed by loosely approximating the measured control system error using the transfer function, $TF_{w\rightarrow e}$. Further discussion of the nuances of this point are provided in [9]. However, everything within the dotted line can be reformulated to represent the closed loop dynamics, where the desired output is $e(t)$. Ultimately, we would like to be able to express these dynamics as shown in Eqns. 74-75 below. In Eqns. 74-75, $x(t)$ is used as a generic placeholder for the transfer function output, $e(t)$, rather than for $x(t)$ shown in Fig. 9. This allows for us to match the notation used in Section IV.

$$\dot{\mathbf{q}}(t) = \mathbf{A}_c\mathbf{q}(t) + \mathbf{B}_w w(t) \quad (74)$$

$$x(t) = \mathbf{C}_c\mathbf{q}(t) + v(t) \quad (75)$$

again, where

$$w(t) \sim \mathcal{N}(0, Q_c)$$
$$v(t) \sim \mathcal{N}(0, R_c)$$

All matrices in the equation above are also subscripted with "c" or "w" in order to disambiguate between the continuous-time dynamics and the discretized dynamics yet to be presented, which mimic those used in Section IV. Eqns. 74-75 need to be discretized in order to fit the digital implementation of the algorithm. However, prior to discretization we can generate statistics from available data that map to parameters in controllable canonical form. The controllable canonical form shown in Eqns. 76-78 includes two intuitive canonical parameters: the natural frequency, ω_n, and the damping ratio, ζ. We can estimate the natural frequency by making an assumption of $x(t)$ to be represented by a zero-mean stationary Gaussian random process. In this case, we can use Rice's formula for the level-upcrossing rate [22], [23], as shown in Eqn. 79, to compute the natural frequency, $\omega_n = \frac{\sigma_{\dot{x}}}{\sigma_x}$. This formula can be derived very easily [24], and is used in similar studies [15], [25], [3].

$$\mathbf{A}_c = \begin{bmatrix} 0 & 1 \\ -\omega_n^2 & -2\zeta\omega_n \end{bmatrix} \quad (76)$$

$$\mathbf{B}_w = \begin{bmatrix} 0 \\ \omega_n^2 \end{bmatrix} \quad (77)$$

$$\mathbf{C}_c = \begin{bmatrix} 1 & 0 \end{bmatrix} \quad (78)$$

$$\nu_x^+ = \frac{\sigma_{\dot{x}}}{2\pi\sigma_x} e^{-\frac{1}{2}\left(\frac{L-\mu_x}{\sigma_x}\right)^2} \quad (79)$$

By using $L = 0$ as a candidate level, we may count the number of zero-upcrossings of the sample data, and compute the 2^{nd}-order statistics, μ_x, and σ_x, in order to use Rice's formula to find ω_n. In case $\mu_x = L = 0$, we simply need ν_x^+, because $\omega_n = \frac{\sigma_{\dot{x}}}{\sigma_x} = 2\pi\nu_x^+$.

After discretization, Eqns. 74-75 fit the modeling paradigm represented by our machine learning problem represented in Fig. 10. The modeling paradigm in this figure is expressed within the probabilistic graphical model framework. Here we can see that the model to be learned is a dynamic system. The observed data are represented by shaded nodes and the unobserved state represented by hidden unshaded nodes. Both are continuous (Gaussian) random variables, the latter of which need to be inferred.

During the learning procedure for the linear dynamic system, the EM algorithm is used to find the parameters shown in Fig. 10. Details of this procedure are provided in Zoubin and Hinton [26] as well as Digalakis et al. [27], and it is implemented using Murphy's BNT (Bayes' Net Toolbox) [28]. Initialization of the parameters shown as θ in Fig. 10 is also performed using some basic heuristics. By initializing $\zeta = 1$ and clamping ω_n during training, we can back out the learned value of the damping ratio ζ. Initial values for $\mathbf{A}_c$ and $\mathbf{B}_w$ can be derived as a function of ζ and ω_n, $\mathbf{C}_c = \begin{bmatrix} 1 & 0 \end{bmatrix}$ is fixed during learning, and R is initialized by making a guess at the SNR (signal to noise ratio), so that $R = \frac{\sigma_x^2}{SNR}$ (σ_x^2 can be computed directly from the data).

Using these assumptions, and by use of steady-state continuous-time Lyapunov equations for Eqns. 74 and 75 (cf. $\mathbf{P}_{ss}^L$ from Fig. 10), we can find an adequate initialization for Q_c, as is performed in [24], [3]. We then discretize all parameters using the sampling interval $T_s = 0.04$ sec (obtained from the data in [9]) , and the procedure outlined in [3], allowing us to form Eqns. 80 - 81, which support the variables shown in Fig. 10. Furthermore, they also support all of the theory presented in Sec. IV, beginning with Eqn. 1, which are identical to Eqns. 80 - 81.

$$\mathbf{q}_{k+1} = \mathbf{A}\mathbf{q}_k + \mathbf{w}_k \quad (80)$$

$$x_k = \mathbf{C}\mathbf{q}_k + v_k \quad (81)$$

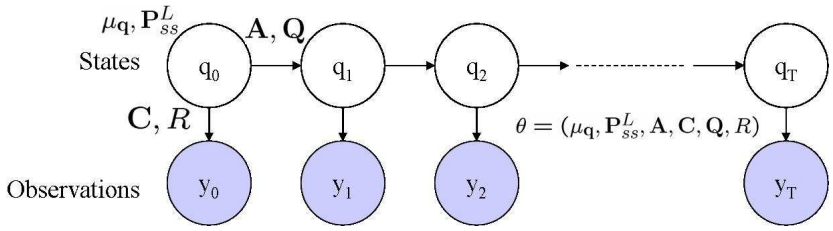

Fig. 10. Linear Dynamic System

where

$$\begin{aligned}
\mathbf{w}_k &\sim \mathcal{N}(0, \mathbf{Q}) \\
v_k &\sim \mathcal{N}(0, R) \\
\mathbf{A} &= e^{\mathbf{A}_c T_s} \\
\mathbf{B} &= (e^{\mathbf{A}_c T_s} - \mathbf{I})\mathbf{A}_c^{-1}\mathbf{B}_w \\
\mathbf{C} &= \mathbf{C}_c \\
R &= \frac{R_c}{T_s} \\
\mathbf{Q} &= \int_0^{T_s} e^{\mathbf{A}_c \lambda}\mathbf{B}_w Q_c \mathbf{B}_w^T e^{\mathbf{A}_c^T \lambda} d\lambda
\end{aligned}$$

Throughout learning, we attempt to retain the controllable canonical structure in order to allow for determination of the learned value for ζ. Furthermore, it allows for an intuitive interpretation of the model's parameters and resulting realizations. This is easily performed by the allowance for enforcement of arbitrary constraints in Murphy's BNT [28], and slight modification of the appropriate open-source routines. Doing so introduces sub-optimality into the learning procedure, which means that the learning curve will not necessarily increase monotonically. However, a reasonable sub-optimal local minimum will be found that best represents the parameter space with enforcement of the controllable canonical form constraint.

In the results section we will perform a comparative analysis of the alarm systems discussed thus far: redline, predictive, and optimal, for each of the level-crossing events introduced in Sec. IV. The manner in which the analysis will be performed is via the ROC curve. For each level-crossing event, four different prediction windows will be investigated: $d = 1$ (0.04 sec), $d = 2$ (0.08 sec), $d = 5$ (0.2 sec), and $d = 10$ (0.4 sec), to allow for the study of a variety of cases for potential early detection. The critical level, L, is chosen for all cases such that $p = 0.001$, by finding an equivalent L_{equiv} using Eqn. 49. It is important to note that for certain level-crossing events, the computational burden of the alarm system design increases with the number of steps in the prediction window.

The data-driven model derived in this section represents a more advanced statistical method than will be presented for the thermal sensation complaint application in the next section. We've demonstrated here that a training procedure involves learning via use of the EM algorithm, etc. Typically, machine learning/data mining methods require a formal experiment to be conducted such that a data set is partitioned into mutually exclusive training and validation hold out subsets. The validation hold out set serves to test the model on data unseen by the model which is derived from empirically-based training data. Some examples of data-driven models that apply this more formal procedure are as follows: [29], [30], [31], [9].

The latter reference uses the same model derived here, where a more formal experimental treatment is provided. We don't use the same formal procedure in this article, because the main point is to introduce an alternative alarm design technique and to compare it to others based upon its theoretical rather than its experimental merits. One of the theoretical merits lies in the fact that we can design an alarm system based upon the model parameters derived from the training data without having to form a ROC curve empirically based upon examples of failures.

B. Thermal Sensation Complaint Application

The second example is based upon a similar state-space model which can be used for complaint prediction in thermal comfort applications. Recall that this model involves a random threshold, and the dynamics are quite different than in the previous example. Here, the hot complaint level is arbitrarily chosen as the example of interest to prevent redundancy. It is assumed that distinct hot and cold optimal alarm systems for the two processes can be designed independently. The state-space systems for the hot critical level as well as the building temperature process of interest can easily be parameterized. The subscripts of the system matrices for both hot complaint and building levels are "h" and "b", respectively, to disambiguate between the continuous time and discretized representations yet to be presented. We first define the state-space equation for the hot complaint level in continuous time:

$$\begin{aligned}
\dot{\mathbf{z}}(t) &= \mathbf{A}_h \mathbf{z}(t) + \mathbf{B}_h n(t) \\
y(t) &= \mathbf{C}_h \mathbf{z}(t)
\end{aligned} \tag{82}$$

The building temperature process is also defined first in continuous time, and both have the same canonical parametrization as in the previous example shown in Eqns. 76-78.

$$\begin{aligned}
\dot{\mathbf{q}}(t) &= \mathbf{A}_b \mathbf{q}(t) + \mathbf{B}_b r(t) + \mathbf{B}_w w(t) \\
x(t) &= \mathbf{C}_b \mathbf{q}(t) + v(t)
\end{aligned} \tag{83}$$

All of these systems must be discretized for the discrete time analysis in Sec. IV to apply. This is done by performing a zero-order hold on all of the above systems as shown in [24]. As such, a sampling interval, T_s, must also be chosen. As a rule

of thumb, we choose a discrete time sampling interval based upon a fraction of the shortest time constant of the dynamics of all relevant processes. More details of the selection of the sampling interval are available in [3]. However, the value of T_s is 20 min, which certainly provides evidence that the dynamics here are quite different than in the previous example. Again, apriori statistics for the zero-mean input and measurement noise processes, $w(t)$, $v(t)$ and $n(t)$ need to be quantified. All are scalar processes, such that $w(t), v(t), n(t) \in \mathbb{R}$. Using Gaussian assumptions, we have the following:

$$\begin{aligned} w(t) &\sim \mathcal{N}(0, Q_w) \\ v(t) &\sim \mathcal{N}(0, R) \\ n(t) &\sim \mathcal{N}(\mu_{T_h}, Q_n) \end{aligned}$$

Note that $r(t) \in \mathbb{R}$ is a scalar fixed control input, which in our application acts as the thermostat setpoint. Note also that R is only applicable to the building temperature process, as it is the measured process, and $n(t)$ is only applicable to the complaint process, where there is no defined "control input." Since we assume that modeling the complaint levels requires no direct control input term, the mean of the input noise driving these processes is taken as the mean of the output. Also, no measurement noise needs to be modeled for the complaint levels because there are no measurements of these unobserved processes. For the building process, typically the output measurements of $x(t)$ come from a DDC (Direct Digital Control) system associated with more sophisticated commercial building systems. Otherwise, these types of measurements often come from micro-dataloggers that record the temperature for a preset period of time.

TABLE I
TABLE OF BUILDING AND COMPLAINT TEMPERATURE STATISTICS

	μ_T	σ_T	$\sigma_{\dot{T}}$	ω	ζ
Hot	91^oF	5.06^oF	$1.14\frac{^oF}{hr}$	$0.6645\frac{rad}{hr}$	1
Cold	50.43^oF	6.14^oF	$4.08\frac{^oF}{hr}$	$0.2253\frac{rad}{hr}$	1
Bldg	74^oF	3.57^oF	$0.91\frac{^oF}{hr}$	$0.2682\frac{rad}{hr}$	1

Table I provides the relevant parameters and respective values for all three processes in continuous time. With the exception of the last two columns, these parameters come from Federspiel's work [25]. The fourth column's (ω) values were derived from the first three using documented methods [15], [25], [24]. The fifth column's (ζ) values were selected heuristically for simulation/analysis ease [24], [32].

The measurement noise can be computed by a procedure discussed in [3], where relevant assumptions are also discussed in detail. The primary reason for computing the measurement noise is for use in Kalman filtering and prediction, which is an implementation prerequisite of the type of optimal alarm system introduced in Sec. IV. The formulae relating the remaining discretized, discrete-time parameters to their continuous-time counterparts are shown in Eqns. 84-87. The state-space parameters are discretized by performing a zero-order hold of the two processes. The input noise discretization results are also provided for the building temperature process, using a documented procedure [33], [34], [35], [36]. The input noise variance for the hot complaint level process was not discretized, but found by using discrete-time Lyapunov analysis of the discretized state-space system. A documented method [24] shows details on how to derive the discrete-time input noise based on a continuous and discrete time Lyapunov analysis of the statistics for these processes.

Discretization of state-space equations for both processes:

$$\mathbf{z}_{k+1} = \mathbf{A}_{hd}\mathbf{z}_k + \mathbf{B}_{hd}n_k \quad (84)$$

$$y_k = \mathbf{C}_{hd}\mathbf{z}_k \quad (85)$$

$$\mathbf{q}_{k+1} = \mathbf{A}\mathbf{q}_k + \mathbf{B}r_k + \mathbf{w}_k \quad (86)$$

$$x_k = \mathbf{C}\mathbf{q}_k + v_k \quad (87)$$

where

$$\begin{aligned} \mathbf{A}_{hd} &= e^{\mathbf{A}_h T_s} \\ \mathbf{B}_{hd} &= (e^{\mathbf{A}_h T_s} - \mathbf{I})\mathbf{A}_h^{-1}\mathbf{B}_h \\ \mathbf{C}_{hd} &= \mathbf{C}_h \\ \mathbf{A} &= e^{\mathbf{A}_b T_s} \\ \mathbf{B} &= (e^{\mathbf{A}_b T_s} - \mathbf{I})\mathbf{A}_b^{-1}\mathbf{B}_b \\ \mathbf{C} &= \mathbf{C}_b \end{aligned}$$

and

$$\begin{aligned} n_k &\sim \mathcal{N}(\mu_T, Q_{n_d}) \\ Q_{n_d} &= E[n_k n_k^T] \\ &= \frac{\sigma_{T_B}^2}{\mathbf{C}_d \tilde{\mathbf{P}}_{ss}^L \mathbf{C}_d^T} \\ \text{where } \tilde{\mathbf{P}}_{ss}^L &\triangleq \frac{\mathbf{P}_{ss}^L}{Q_{n_d}} \\ &= \mathbf{A}_d \tilde{\mathbf{P}}_{ss}^L \mathbf{A}_d^T + \mathbf{B}_{w_d}\mathbf{B}_{w_d}^T \end{aligned}$$

also:

$$\begin{aligned} \mathbf{w}_k &\sim \mathcal{N}(\mathbf{0}, \mathbf{Q}) \\ \mathbf{Q} &= \int_0^{T_s} e^{\mathbf{A}\lambda}\mathbf{B}_w Q_w \mathbf{B}_w^T e^{\mathbf{A}^T\lambda} d\lambda \\ \text{where } Q_w &= \frac{\sigma_{T_B}^2}{\mathbf{C}\tilde{\mathbf{P}}_c\mathbf{C}^T} \\ \text{and } \tilde{\mathbf{P}}_c &\triangleq \frac{\mathbf{P}_c}{Q_w} \\ \mathbf{A}\tilde{\mathbf{P}}_c + \tilde{\mathbf{P}}_c\mathbf{A}^T &= \mathbf{B}_w\mathbf{B}_w^T \end{aligned}$$

and:

$$\begin{aligned} v_k &\sim \mathcal{N}(0, R) \\ R &= E[v_k v_k^T] \end{aligned}$$

Equations 84-87 relate back to Eqn. 1 in support of theory presented in Sec. IV. This is not as apparent as in the previous example, where there was a more straightforward application of the theory developed. In this case, the level-crossing problem can be reformulated because two processes interact, the stochastic critical level associated with hot complaints, and the building temperature process. In this case the Kalman filtering and prediction would only be performed on the building temperature process. Reformulation of the level-crossing problem entails transformation of the problem into one that fits the paradigm of a fixed, static threshold. In doing so we simply take the difference between the stochastic critical level, y_k, and the controlled process, x_k, implying that the critical level is given by $L = 0$, due to upcrossings of $L = 0$ by $x_k - y_k$.

Furthermore, the conditional expectations necessary for arriving at the Kalman filtering and prediction formulae require a slight recasting. This will not result in a large deviation from the alarm theory previously presented, but the $\mathbf{B}r_k$ term in Eqn. 86 needs to be accounted for and propagated through all of the equations in Sec. IV. In addition, the level-crossing problem can be reformulated even further due to the fact that the conditional expectation of the stochastic critical level, y_k, conditioned on the observations is given as shown in Eqns. 88-90. This is due to the assumption that building temperatures are uncorrelated with, or independent of the hot and cold complaint levels. This independence assumption made in Federspiel's study [14], is valid if the coping behaviors of building occupants do not vary with building temperature. As such, the reformulation of the problem from the alarm theory standpoint as presented in Sec. IV is for the process, x_k, to interact via upcrossing with the level $L = \mu_{T_h}$.

$$\begin{aligned} \hat{y}_{k+d|k} &\triangleq E[y_{k+d}|x_0, \ldots, x_k] && (88) \\ &= \hat{y}_{k+i|k}, \forall i \geq 0 && (89) \\ &= \mu_{T_h} && (90) \end{aligned}$$

Again, our problem is to design an optimal alarm system that predicts at least one operating complaint, or at least one arrival and operating complaint. For the former case, we use a 12-step ahead prediction window, or $m = 12$. This corresponds to a 4-hour period in which operating complaints can occur either during the morning or afternoon period of the day. For the latter case, we use the same $m = 12$ step ahead prediction window, and $d = 3$ as the number of steps prior to the start of the day. In this 1-hour period prior to the beginning of the day, we want to predict both arrival complaints and operating complaints in the ensuing 4-hour period.

All results and subsequent discussion for the application of the theory presented in Sec. IV to this example are provided in thorough detail in [3], in lieu of presentation in this article. In [3], a comparative analysis of the differing approximations used for the level crossing events defined above is provided. Furthermore, an error analysis for these approximations as well as comments on accuracy and computational design time are provided. Unlike the previous application, the alarm system metrics used are Type II error (missed detection) and false alarm probabilities in lieu of the ROC curve.

The extent of the data-driven method as applied to this example lies in generation of a model derived from statistics. It is still possible to characterize the building temperature process through the same control loop as shown in Fig. 9 as for the previous example, although there is a different transfer function of interest, $TF_{r \to x}$. However, the parameters shown in Fig. 10 are not learned via the EM algorithm for this example. Rather, the model derived purely from statistics that would typically serve as the starting point for such an algorithm is used. However, if we apply the canonical constraints as for the previous example, the resulting distribution based on the learned parameters may not change significantly from the initial parameters. As such, using the model derived from statistics is a feasible alternative for this application.

VI. Results

All of the results presented in this section are only for the spacecraft propulsion anomaly detection example. As stated previously, detailed discussion of the results for the thermal sensation complaint example can be found in [3]. Chapter 3 of [3] details the fidelity of various approximations to the exact optimal alarm region, and Chapter 7 covers the implementation of the resulting alarm systems. In general it was found that improved approximations to the exact alarm region were possible at the expense of an increased computational burden.

There are no comparisons to other types of alarm systems in [3] as will be presented here. As such, we present the results for all alarm system types and each of the level-crossing events covered in Sec. IV. ROC curves provide the basis for Figs. 11-15 presented in this section. The formulae for true and false positive rates required to form the ROC curves were presented in Sec. IV as Eqns. 47-48. We know that as long as the following three probability computations are performed: $P(C_{exact})$, $P(A_{approx})$, and $P(C_{exact}, A_{approx})$, any relevant alarm system metric can be derived from them, including the true and false positive rates required to form the ROC curves.

The formulae for each of these probabilities were provided in detail for all alarm system types and level-crossing events of interest, both in Sec. IV and in the appendices. As such, computing integrals of the form necessary for design and comparison of the alarm systems require multivariate probability computations. These computations are performed by using Genz's algorithm [17], based upon a Monte Carlo-style integration. Due to the Monte Carlo nature of the computations, a fixed number of random samples must be set. For the results presented below, the number of random samples for each integration performed varied between 360 and 36000 sample points, depending on the resolution required for sufficient smoothness to be attained.

There is an important property of the ROC curve to consider when evaluating the results presented in the section from an absolute standpoint. The diagonal line corresponding to equivalent true and false positives values represents the boundary

above which a system performs better than randomly guessing if the level-crossing event occurs. From a relative standpoint, this property is less important because the objective is to perform a qualitative comparison among different types of alarm systems.

Many of the results shown in Figs. 11-15 actually lie very close to, if not along the random guessing line. This relates to the model fidelity (it's capability of explaining the data used to train it), the chosen threshold, L_{equiv}, and the prediction window size. For extreme values of the threshold that correspond to $p = 0.001$, and for prediction windows that span a large number of time steps (i.e. $d = 10$), there is naturally more difficultly in accurately predicting events. Furthermore, the ROC curve is only as good as the given model parameters, since the ROC statistics are a function of these model parameters, which are implicitly a function of the data. Therefore, a different training data set may have resulted in a ROC curve that has a different shape that would be more robust to increases in the size of the prediction window.

For the curves presented in Figs. 11-15 below, the results of the optimal ROC curve true and false positive rates were obtained via both approximations studied. As presented in Sec. IV, the two optimal alarm region approximations studied were the "closed-form" and the "root-finding" approximations. Both approximations yielded very similar and in some cases nearly indistinguishable results for all of the cases presented below. However, both approximations are displayed in Figs. 11-15 below, with "CF" corresponding to "closed-form," and "RF" corresponding to "root-finding" in the legend.

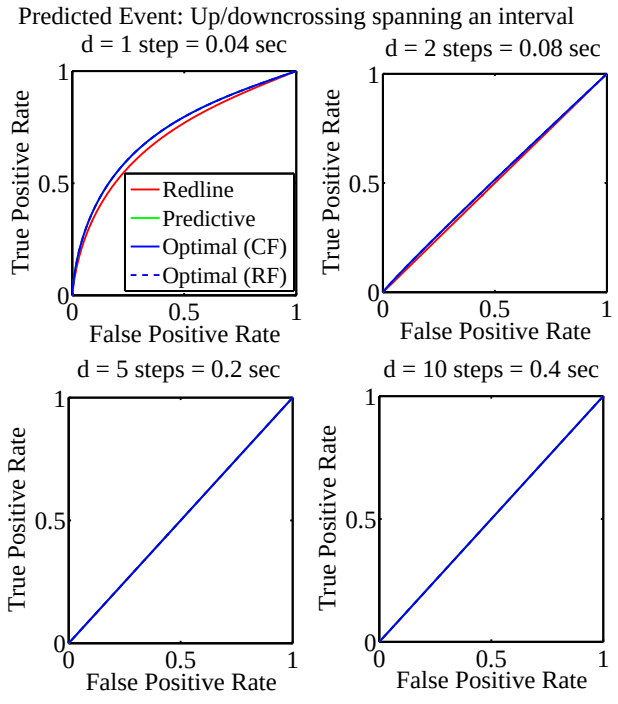

Fig. 11. ROC curve for up/downcrossing event spanning an interval

A. Up/downcrossing event spanning an interval

By observing the ROC curves in Fig. 11, we can immediately discern that for the two smallest prediction windows ($d = 1$, $d = 2$), both optimal alarm system approximations and the predictive alarm system perform better than the redline alarm system, and their performance appears qualitatively identical (they are superimposed). As the prediction window increases, all alarm systems appear to have identical performance, which lie along the random guessing line.

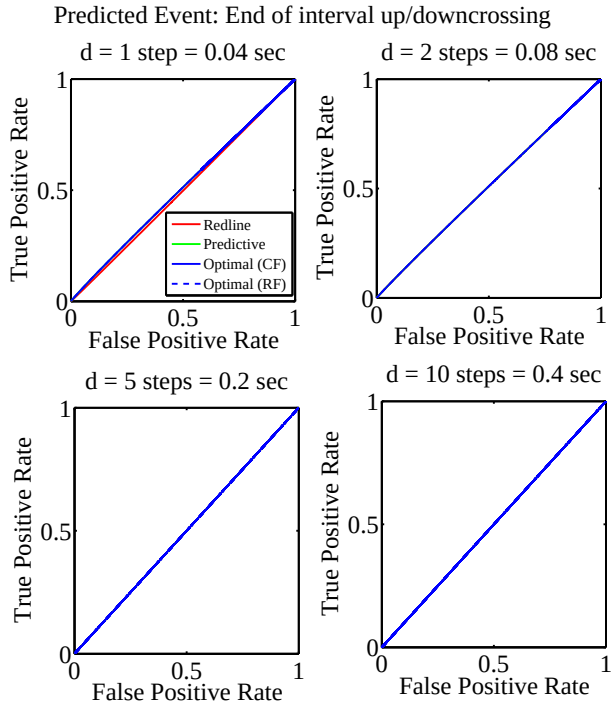

Fig. 12. ROC curve for end of interval up/downcrossing event

B. End of interval up/downcrossing event

By observing the ROC curves in Fig. 12, we can tell that for the smallest prediction window ($d = 1$), both optimal alarm system approximations and the predictive alarm system perform better than the redline alarm system, and their performance again appears qualitatively identical. Not much else can be concluded for the remaining prediction windows because they yield performance that lies along the random guessing line. This can mainly be attributed to the fact that the up/downcrossing event occurs at the end of the prediction window interval, making it a difficult event to predict, and thus a poor candidate for an alarm system.

C. End of interval exceedance/fade event

In Fig. 13 we can qualitatively observe that the performance of all alarm systems is nearly identical for all prediction windows to that of Fig. 11, for the up/downcrossing event spanning an interval. This may indicate that these two events are equally viable candidates for developing alarm systems.

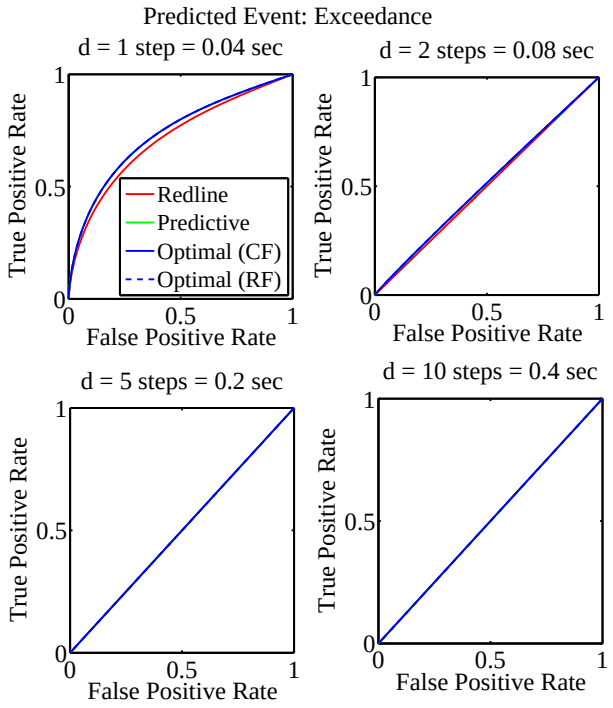

Fig. 13. ROC curve for end of interval exceedance/fade event

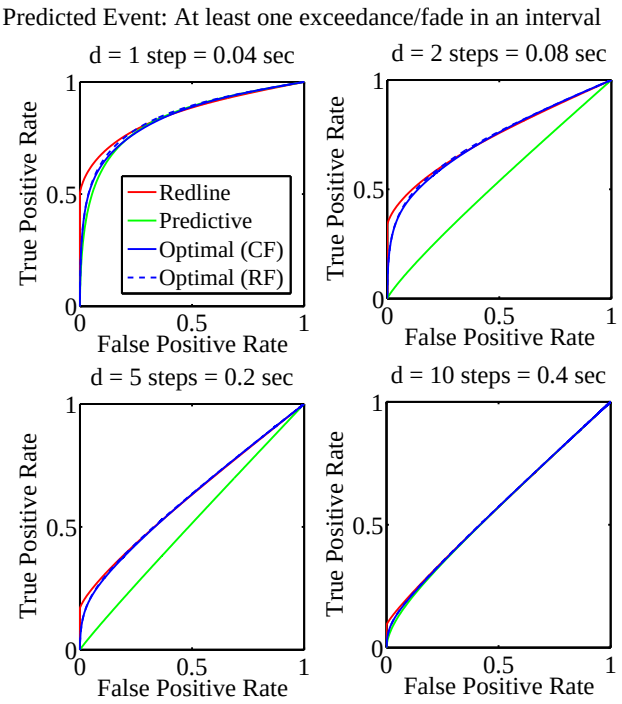

Fig. 14. ROC curve for at least one exceedance/fade event within an interval

D. At least one exceedance/fade event within an interval

Fig. 14 exhibits the first characteristics of a discernable distinction between the two different approximations to the optimal alarm system, specifically for the smallest prediction window ($d = 1$). We can tell that the "root-finding" approximation to the optimal alarm system outperforms the "closed-form" approximation and the predictive alarm system in a qualitative sense, for all values of P_b. In fact, although it is difficult to discern by observing the qualitative nature of the plot, the "closed-form" approximation outperforms the predictive alarm system only for a limited set of values for P_b.

Most interestingly, none of the alarm systems consistently outperforms the redline alarm system for all values of P_b. However, each of these alarm systems does outperform the redline alarm system over various ranges of P_b values. The range of values is largest for the "root-finding" approximation to the optimal alarm system, followed by the predictive alarm system, and is least for the "closed-form" approximation to the optimal alarm system. As the prediction window size increases, it becomes clear the both "closed-form" and "root-finding" approximations to the optimal alarm system outperform the predictive alarm system, whose ROC curve approaches the random guess line cyclically as d increases. However, neither still consistently outperforms the redline alarm system for all values of P_b for these larger prediction windows. It therefore may be of use to study this event further, and to use the area under the ROC curve to obtain a comparison that is independent of the P_b value.

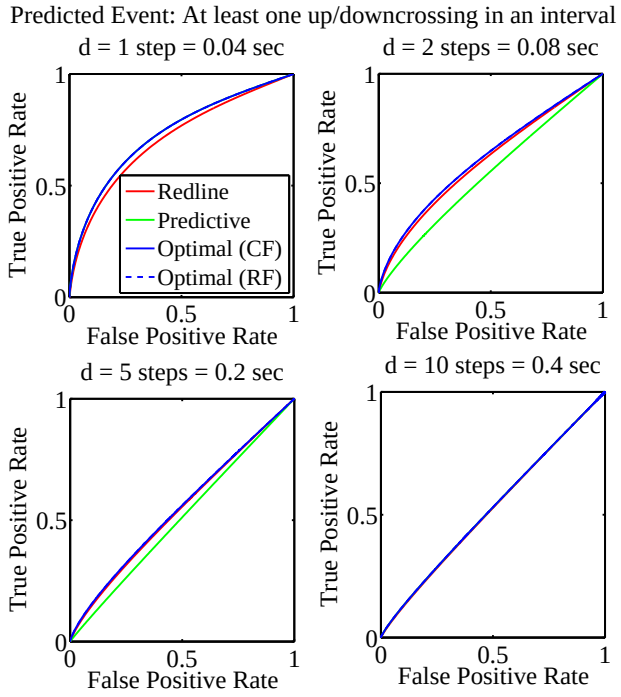

Fig. 15. ROC curve for at least one up/downcrossing event within an interval

E. At least one up/downcrossing event within an interval

Finally, in Fig. 15 we can observe that the performance of both approximations to the optimal alarm system is always better than both the redline and predictive alarm system for all prediction windows. This is most evident for all but the largest prediction window of d=10. Curiously, again as the

prediction window increases, the predictive alarm system ROC curve approaches the random guess line in a cyclical manner, as distinct from all other alarm systems which approach the random guess line asymptotically as d increases.

VII. Discussion, Conclusions, & Future Research

It is not easy to draw a general conclusion from the results presented in the previous section. Comparatively, in consideration of all alarm systems, determining the best performer depends on the type of level-crossing event. However in most cases, the optimal alarm system approximations are clearly the better performers, with the event that characterizes at least one exceedance/fade event within an interval as the exception. In this case, the redline alarm system outperforms the others over a fixed range of threshold values. Even so, we have gained invaluable insight into which events are difficult to predict, and thus poor candidates for alarm systems, and which events are prime candidates for further study.

As mentioned previously, the results are also affected by extreme values of the chosen threshold, and for prediction windows that span a large number of time steps. For another application with a different prediction window, critical threshold, and dynamics, the results of the optimal alarm system may yield better performance than the redline alarm system. It is also possible a better approximation to the optimal alarm region might also close the optimality gap in this particular instance. Furthermore, the use of Kalman filtering allows for estimation and prediction of states that would otherwise be unobserved or immeasurable using the redline alarm system.

For the spacecraft propulsion anomaly detection example, two distinct levels were used for redline analysis: one for limit checking, and one for alarm design. This technique was used in lieu of a single level having both functions for the redline analysis because the ROC statistics can be expressed as a function of the model parameters when using two levels. This is not always possible for analysis in the case of a single level, specifically in regards to the technique used to generate the ROC curve statistics. Furthermore, two levels are often used in practice for the design of fault detection algorithms that involve limit-based abort decisions. A "yellow-line" limit check is often used as a precursor caution and warning threshold to the "redline" abort threshold. The former can be used as an alarm system design parameter, where the latter may serve as a hard limit determined apriori via extensive experimental validation.

When there are insufficient examples of failures, the ROC curve statistics (the true and false positive rates) can be estimated empirically as limiting fractions, using only a single redline level. This is akin to the "counting" method discussed in the introduction, and as such it might also be possible to simulate failures with the model in lieu of using actual observations of failures. As such, this method could have been used to demonstrate a comparative analysis of a redline alarm system, based upon a single level. This "Monte-Carlo" style technique is computationally intensive, and is still based upon the model-generated failures as opposed to actual observations of failures. A similar empirical approach detailed in [20] can be used to form the ROC curve.

Alternatively, we could have used the model parameters using the χ^2 distribution as presented in Sec. IV for the analysis involving a single level. However, this technique will generate ROC curve statistics that belong to two complementary hypotheses. This paradigm differs from the one used to generate ROC curve statistics presented throughout this paper. We based our alarm statistics upon distinct definitions of a critical event as the hypothesis, and an independently designed alarm system. As such, ROC analysis cannot be performed using our paradigm in the case of a single redline level. However, in the other paradigm, alarm systems merge the functionality of limit checking and the use of an alarm design parameter into a test of two complementary hypotheses.

Using this method, it is not possible to decouple independent alarm design from the critical event, which provides a measure of functional distinction. This method is also the one most commonly found in the literature, i.e., [20], [10], [11]. Arguably, the critical event should be based upon the physics of the failure, and the alarm design parameter should be used to predict it. The distinction between these two paradigms is one of the most discernable differences in the theoretical techniques used here and in other literature, [1], [2], [3], [4], [5], [6].

Subject to certain constraints, design of the alarm system can proceed without the need to observe actual examples of failures, and there is no need to estimate the alarm system metrics empirically using either paradigm. This obviates the need to rely upon having actual available examples of failures for alarm system design to generate the ROC curve. That is because they are based on the model and design parameters. However, the hypothesis-based level-crossing event must sufficiently characterize an actual physical failure for the model-based analysis to be of great benefit.

All of the alarm theory presented in this paper has also been supported by the thermal sensation complaint application. The resulting details are presented in [3]. We present both examples as means to motivate the use of such algorithmic novelty for other potential applications that require health management or fault detection. The basis of the theory itself is quite pedantic, and can often be found in pedagogical introductions to modern control theory. It can therefore naturally be extended to more realistic scenarios that include non-linear dynamic systems, non-Gaussian distributions, and potentially the use of extended/unscented Kalman filtering, and/or particle filtering. In addition, adaptive model updates may be considered, as is in work presented by Antunes et al. [8].

Nothing yet has been mentioned about the actual alarm design procedure or results from its implementation, which can be performed by selecting the optimal border probability, P_b. This border probability serves as a free parameter, and hence as the primary design metric. The steps required for design of an alarm system have been covered in previous work [1], [2]. However, they tend to be based either on purely heuristic trial-and-error approaches, or cost functions. For cost functions, sometimes it is easy to assign particular costs to events that penalize the probability of alarms, false alarms, and missed detections, etc. Assigning these costs requires heuristic knowledge of the risk-reward tradeoff in terms of relevant

alarm system metrics. This is cause for further study, in which these challenging design and implementation issues will be covered in earnest. In future work we can look at creative ways to select the value of P_b for the optimal alarm system given the appropriate criteria and investigating different cost functions.

An issue which has not been addressed in detail for the spacecraft propulsion anomaly detection application is the quantification of the approximations used. In [3], the error introduced in such approximations was discussed in detail for the thermal sensation complaint application. We would like to be able to determine: (1) the level of sub-optimality introduced by making these approximations for the spacecraft propulsion anomaly detection application, and (2) if they can be improved. Such issues may relate to why we found an improbable exception in our results indicating superior performance of the redline alarm system. It is known that better approximations to the two-dimensional alarm regions are possible, documented in Chap. 4 of [3]. Extensions to these same sorts of approximations to alarm regions that "shrink wrap" the exact alarm region under certain conditions are cause for further investigation for alarm regions in multiple dimensions.

The basic engineering approximations introduced in this article addressed the main objective of demonstrating the ability to initiate an alarm with as large a prediction window as possible in advance of critical events. Our presentation here acts as a necessary precursor to the computationally efficient design and implementation of optimal alarm systems, as well as improvements to these approximations to be presented in sequel articles. Furthermore, the theoretical novelty of this paper has been demonstrated, in an aim to participate in the Kalman filter-based fault detection literature discussion from a different theoretical angle. In doing so, we hope to have more precisely closed the gap between the use of Kalman prediction techniques and optimal alarm systems.

APPENDIX I
REDLINE ALARM SYSTEM COMPUTATIONS FOR $P(C_{exact}, A_{approx})$

A. Up/downcrossing event spanning an interval

The probability, $P(C_{exact}, A_{approx})$, for this event is shown in Eqn. I.1.

$$P(C_{exact}, A_{approx}) =$$

$$\begin{aligned}
&\begin{cases} 0 & L_A > L \\ P(L_A < x_k < L, |x_{k+d}| > L) + \dots \\ P(-L < x_k < -L_A, |x_{k+d}| > L) & L_A < L \end{cases} \\
&= \begin{cases} 0 & L_A > L \\ \int_L^\infty \int_{L_A}^L \mathcal{N}(\mathbf{x}; \mu_\mathbf{x}, \Sigma_\mathbf{x})\ d\mathbf{x} + \dots \\ \int_{-\infty}^{-L} \int_{L_A}^L \mathcal{N}(\mathbf{x}; \mu_\mathbf{x}, \Sigma_\mathbf{x})\ d\mathbf{x} + \dots & L_A < L \\ \int_L^\infty \int_{-L}^{-L_A} \mathcal{N}(\mathbf{x}; \mu_\mathbf{x}, \Sigma_\mathbf{x})\ d\mathbf{x} + \dots \\ \int_{-\infty}^{-L} \int_{-L}^{-L_A} \mathcal{N}(\mathbf{x}; \mu_\mathbf{x}, \Sigma_\mathbf{x})\ d\mathbf{x} \end{cases} \quad \text{(I.1)} \\
\mathbf{x} &= \begin{bmatrix} x_k \\ x_{k+d} \end{bmatrix} \in \mathbb{R}^2
\end{aligned}$$

B. End of interval up/downcrossing event

The probability, $P(C_{exact}, A_{approx})$, for this event is shown in Eqn. I.2.

$$P(C_{exact}, A_{approx}) =$$

$$\begin{aligned}
&= P(|x_k| > L_A, |x_{k+d}| < L, |x_{k+d+1}| > L) \\
&= \int_L^\infty \int_{-L}^L \int_{L_A}^\infty \mathcal{N}(\mathbf{x}; \mu_\mathbf{x}, \Sigma_\mathbf{x})\ d\mathbf{x} + \dots \\
&\quad \int_{-\infty}^{-L} \int_{-L}^L \int_{L_A}^\infty \mathcal{N}(\mathbf{x}; \mu_\mathbf{x}, \Sigma_\mathbf{x})\ d\mathbf{x} + \dots \\
&\quad \int_L^\infty \int_{-L}^L \int_{-\infty}^{-L} \mathcal{N}(\mathbf{x}; \mu_\mathbf{x}, \Sigma_\mathbf{x})\ d\mathbf{x} + \dots \\
&\quad \int_{-\infty}^{-L} \int_{-L}^L \int_{-\infty}^{-L} \mathcal{N}(\mathbf{x}; \mu_\mathbf{x}, \Sigma_\mathbf{x})\ d\mathbf{x} \quad \text{(I.2)} \\
\mathbf{x} &= \begin{bmatrix} x_k \\ x_{k+d} \\ x_{k+d+1} \end{bmatrix} \in \mathbb{R}^3
\end{aligned}$$

C. End of interval exceedance/fade event

The probability, $P(C_{exact}, A_{approx})$, for this event is shown in Eqn. I.3.

$$P(C_{exact}, A_{approx}) =$$

$$\begin{cases} P(|x_k| > \max(L, L_A)) & d = 0 \\ P(|x_k| > L_A, |x_{k+d}| > L) & d > 0 \end{cases} \quad \text{(I.3)}$$

where $P(|x_k| > \max(L, L_A)) = 2\Phi\left(-\frac{\max(L, L_A)}{\sqrt{\mathbf{C}\mathbf{P}_{ss}^L\mathbf{C}^T + R}}\right)$

and $P(|x_k| > L_A, |x_{k+d}| > L) =$

$$\begin{aligned}
&\int_L^\infty \int_{L_A}^L \mathcal{N}(\mathbf{x}; \mu_\mathbf{x}, \Sigma_\mathbf{x})\ d\mathbf{x} + \dots \\
&\int_{-\infty}^{-L} \int_{L_A}^L \mathcal{N}(\mathbf{x}; \mu_\mathbf{x}, \Sigma_\mathbf{x})\ d\mathbf{x} + \dots \\
&\int_L^\infty \int_{-L}^{-L_A} \mathcal{N}(\mathbf{x}; \mu_\mathbf{x}, \Sigma_\mathbf{x})\ d\mathbf{x} + \dots \\
&\int_{-\infty}^{-L} \int_{-L}^{-L_A} \mathcal{N}(\mathbf{x}; \mu_\mathbf{x}, \Sigma_\mathbf{x})\ d\mathbf{x} \\
\mathbf{x} &= \begin{bmatrix} x_k \\ x_{k+d} \end{bmatrix} \in \mathbb{R}^2
\end{aligned}$$

D. At least one exceedance/fade event within an interval

The probability, $P(C_{exact}, A_{approx})$, for this event is shown in Eqn. I.4.

$$P(C_{exact}, A_{approx}) =$$

$$\begin{cases} P(|x_k| > \max(L, L_A)) + \dots & L_A < L \\ \sum_{j=1}^d P_{a_j}^+ + P_{a_j}^- & \\ P(|x_k| > \max(L, L_A)) & L_A > L \end{cases} \quad \text{(I.4)}$$

where

$$
\begin{aligned}
P_{a_j}^{+} &= P(L_A < x_k < L, \bigcap_{i=1}^{j-1} |x_{k+i}| < L, |x_{k+j}| > L) \\
&= \int_{L}^{\infty}\int_{-L}^{L}\cdots\int_{-L}^{L}\int_{L_A}^{L} \mathcal{N}(\mathbf{x}_j;\mu_{\mathbf{x}_j},\Sigma_{\mathbf{x}_j})\ d\mathbf{x}_j + \ldots \\
&\quad \underbrace{\int_{-\infty}^{-L}\int_{-L}^{L}\cdots\int_{-L}^{L}\int_{L_A}^{L} \mathcal{N}(\mathbf{x}_j;\mu_{\mathbf{x}_j},\Sigma_{\mathbf{x}_j})\ d\mathbf{x}_j}_{j+1} \\
P_{a_j}^{-} &= P(-L < x_k < -L_A, \bigcap_{i=1}^{j-1} |x_{k+i}| < L, |x_{k+j}| > L) \\
&= \int_{L}^{\infty}\int_{-L}^{L}\cdots\int_{-L}^{L}\int_{-L}^{-L_A} \mathcal{N}(\mathbf{x}_j;\mu_{\mathbf{x}_j},\Sigma_{\mathbf{x}_j})\ d\mathbf{x}_j + \ldots \\
&\quad \underbrace{\int_{-\infty}^{-L}\int_{-L}^{L}\cdots\int_{-L}^{L}\int_{-L}^{-L_A} \mathcal{N}(\mathbf{x}_j;\mu_{\mathbf{x}_j},\Sigma_{\mathbf{x}_j})\ d\mathbf{x}_j}_{j+1} \\
\mathbf{x}_j &= \begin{bmatrix} x_k \\ \vdots \\ x_{k+j} \end{bmatrix} \in \mathbb{R}^{j+1}
\end{aligned}
$$

E. At least one up/downcrossing event within an interval

Finally, the probability, $P(C_{exact}, A_{approx})$, for this event is shown in Eqn. I.5, with identical definitions of $P_{a_j}^{+}$,$P_{a_j}^{-}$, and $\mathbf{x}_j$ as in the previous case.

$$
P(C_{exact}, A_{approx}) = \begin{cases} \sum_{j=1}^{d} P_{a_j}^{+} + P_{a_j}^{-} & L_A < L \\ 0 & L_A > L \end{cases} \quad \text{(I.5)}
$$

APPENDIX II
PREDICTIVE ALARM SYSTEM COMPUTATIONS FOR $P(C_{exact}, A_{approx})$

A. Up/downcrossing event spanning an interval

We begin with the probability, $P(C_{exact}, A_{approx})$, for this event, shown in Eqn. II.1.

$P(C_{exact}, A_{approx}) =$

$$
\begin{aligned}
&= P(|x_k| < L, |x_{k+d}| > L, |\hat{x}_{k+d|k}| > L_A) \\
&= \int_{L_A}^{\infty}\int_{L}^{\infty}\int_{-L}^{L} \mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})\ d\mathbf{x} + \ldots \\
&\quad \int_{-\infty}^{-L_A}\int_{L}^{\infty}\int_{-L}^{L} \mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})\ d\mathbf{x} + \ldots \\
&\quad \int_{L_A}^{\infty}\int_{-\infty}^{-L}\int_{-L}^{L} \mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})\ d\mathbf{x} + \ldots \\
&\quad \int_{-\infty}^{-L_A}\int_{-\infty}^{-L}\int_{-L}^{L} \mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})\ d\mathbf{x} \qquad \text{(II.1)} \\
\mathbf{x} &= \begin{bmatrix} x_k \\ x_{k+d} \\ \hat{x}_{k+d|k} \end{bmatrix}, \mu_{\mathbf{x}} = \begin{bmatrix} \mathbf{C}\mu_{\mathbf{q}} \\ \mathbf{C}\mu_{\mathbf{q}} \\ \mathbf{C}\mu_{\mathbf{q}} \end{bmatrix} \in \mathbb{R}^3
\end{aligned}
$$

$\Sigma_{\mathbf{x}_j}(i_1, i_2) =$

$$
\begin{cases} \mathbf{C}\mathbf{P}_{ss}^{L}\mathbf{C}^T + \mathbf{R} & 1 \le i_1 = i_2 \le 2 \\ \mathbf{C}\mathbf{A}^d(\mathbf{P}_{ss}^{L} - \hat{\mathbf{P}}_{ss}^{R})(\mathbf{A}^d)^T\mathbf{C}^T & i_1 = i_2 = 3 \\ \mathbf{C}\mathbf{P}_{ss}^{L}(\mathbf{A}^d)^T\mathbf{C}^T & 1 \le i_1 \ne i_2 \le 2 \\ \mathbf{C}\mathbf{A}^d(\mathbf{P}_{ss}^{L} - \hat{\mathbf{P}}_{ss}^{R})\mathbf{C}^T & i_1 \ne i_2 : i_1 \vee i_2 = 3 \end{cases}
$$

B. End of interval up/downcrossing event

The probability, $P(C_{exact}, A_{approx})$, for this event is shown in Eqn. II.2.

$P(C_{exact}, A_{approx}) =$

$$
\begin{aligned}
&= P(|x_{k+d}| < L, |x_{k+d+1}| > L, |\hat{x}_{k+d+1|k}| > L_A) \\
&= \int_{L_A}^{\infty}\int_{L}^{\infty}\int_{-L}^{L} \mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})\ d\mathbf{x} + \ldots \\
&\quad \int_{-\infty}^{-L_A}\int_{L}^{\infty}\int_{-L}^{L} \mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})\ d\mathbf{x} + \ldots \\
&\quad \int_{L_A}^{\infty}\int_{-\infty}^{-L}\int_{-L}^{L} \mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})\ d\mathbf{x} + \ldots \\
&\quad \int_{-\infty}^{-L_A}\int_{-\infty}^{-L}\int_{-L}^{L} \mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})\ d\mathbf{x} \qquad \text{(II.2)} \\
\mathbf{x} &= \begin{bmatrix} x_{k+d} \\ x_{k+d+1} \\ \hat{x}_{k+d+1|k} \end{bmatrix}, \mu_{\mathbf{x}} = \begin{bmatrix} \mathbf{C}\mu_{\mathbf{q}} \\ \mathbf{C}\mu_{\mathbf{q}} \\ \mathbf{C}\mu_{\mathbf{q}} \end{bmatrix} \in \mathbb{R}^3
\end{aligned}
$$

$\Sigma_{\mathbf{x}_j}(i_1, i_2) =$

$$
\begin{cases} \mathbf{C}\mathbf{P}_{ss}^{L}\mathbf{C}^T + \mathbf{R} & 1 \le i_1 = i_2 \le 2 \\ \mathbf{C}\mathbf{A}^{d+1}(\mathbf{P}_{ss}^{L} - \hat{\mathbf{P}}_{ss}^{R})(\mathbf{A}^{d+1})^T\mathbf{C}^T & i_1 = i_2 = 3 \\ \mathbf{C}\mathbf{X}_{ss}\mathbf{C}^T & 1 \le i_1 \ne i_2 \le 2 \\ \mathbf{C}\mathbf{A}^d(\mathbf{P}_{ss}^{L} - \hat{\mathbf{P}}_{ss}^{R})(\mathbf{A}^{d+1})^T\mathbf{C}^T & i_1 \ne i_2 : i_1 \vee i_2 = 3 \end{cases}
$$

$$
\begin{aligned}
\mathbf{X}_{ss} &= \mathbf{A}^d\mathbf{P}_{ss}^{L}(\mathbf{A}^{d+1})^T + \mathbf{L}_{ss} - \mathbf{A}^d\mathbf{L}_{ss}(\mathbf{A}^d)^T \\
\mathbf{L}_{ss} &= \mathbf{A}\mathbf{L}_{ss}\mathbf{A}^T + \mathbf{A}\mathbf{Q}
\end{aligned}
$$

C. End of interval exceedance/fade event

The probability, $P(C_{exact}, A_{approx})$, for this event is shown in Eqn. II.3.

$P(C_{exact}, A_{approx}) =$

$$
\begin{aligned}
&P(|x_{k+d}| > L, |\hat{x}_{k+d|k}| > L_A) \\
&\int_{L_A}^{\infty}\int_{L}^{\infty} \mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})\ d\mathbf{x} + \ldots \\
&\int_{-\infty}^{-L_A}\int_{L}^{\infty} \mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})\ d\mathbf{x} + \ldots \\
&\int_{L_A}^{\infty}\int_{-\infty}^{-L} \mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})\ d\mathbf{x} + \ldots \\
&\int_{-\infty}^{-L_A}\int_{-\infty}^{-L} \mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})\ d\mathbf{x} \qquad \text{(II.3)} \\
\mathbf{x} &= \begin{bmatrix} x_{k+d} \\ \hat{x}_{k+d|k} \end{bmatrix} \in \mathbb{R}^2
\end{aligned}
$$

$$
\Sigma_{\mathbf{x}_j}(i_1, i_2) = \begin{cases} \mathbf{C}\mathbf{P}_{ss}^{L}\mathbf{C}^T + \mathbf{R} & i_1 = i_2 = 1 \\ \mathbf{C}\mathbf{A}^d(\mathbf{P}_{ss}^{L} - \hat{\mathbf{P}}_{ss}^{R})(\mathbf{A}^d)^T\mathbf{C}^T & \text{o.w.} \end{cases}
$$

D. At least one exceedance/fade event within an interval

The probability, $P(C_{exact}, A_{approx})$, for this event is shown in Eqn. II.4.

$P(C_{exact}, A_{approx}) =$

$$P(|x_k| > L, |\hat{x}_{k+d|k}| > L_A) + \sum_{j=1}^{d} P_{a_j} \qquad \text{(II.4)}$$

where $P(|x_k| > L, |\hat{x}_{k+d|k}| > L_A) =$

$$\int_{L_A}^{\infty}\int_{L}^{\infty} \mathcal{N}(\mathbf{x}; \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})\ d\mathbf{x} + \ldots$$
$$\int_{-\infty}^{-L_A}\int_{L}^{\infty} \mathcal{N}(\mathbf{x}; \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})\ d\mathbf{x} + \ldots$$
$$\int_{L_A}^{\infty}\int_{-\infty}^{-L} \mathcal{N}(\mathbf{x}; \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})\ d\mathbf{x} + \ldots$$
$$\int_{-\infty}^{-L_A}\int_{-\infty}^{-L} \mathcal{N}(\mathbf{x}; \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})\ d\mathbf{x}$$

$$\mathbf{x} = \begin{bmatrix} x_k \\ \hat{x}_{k+d|k} \end{bmatrix} \in \mathbb{R}^2$$

$$\Sigma_{\mathbf{x}_j}(i_1, i_2) = \begin{cases} \mathbf{C}\mathbf{P}_{ss}^{L}\mathbf{C}^T + \mathbf{R} & i_1 = i_2 = 1 \\ \mathbf{C}\mathbf{A}^d(\mathbf{P}_{ss}^{L} - \hat{\mathbf{P}}_{ss}^{R})(\mathbf{A}^d)^T\mathbf{C}^T & i_1 = i_2 = 2 \\ \mathbf{C}(\mathbf{P}_{ss}^{L} - \hat{\mathbf{P}}_{ss}^{R})(\mathbf{A}^d)^T\mathbf{C}^T & i_1 \neq i_2 \end{cases}$$

and

$$P_{a_j} = P(\bigcap_{i=0}^{j-1} |x_{k+i}| < L, |x_{k+j}| > L, \hat{x}_{k+d|k} > L_A)$$
$$= \int_{L_A}^{\infty}\int_{L}^{\infty}\int_{-L}^{L}\cdots\int_{-L}^{L} \mathcal{N}(\mathbf{x}_j; \mu_{\mathbf{x}_j}, \Sigma_{\mathbf{x}_j})\ d\mathbf{x}_j + \ldots$$
$$\int_{L_A}^{\infty}\int_{-\infty}^{-L}\int_{-L}^{L}\cdots\int_{-L}^{L} \mathcal{N}(\mathbf{x}_j; \mu_{\mathbf{x}_j}, \Sigma_{\mathbf{x}_j})\ d\mathbf{x}_j + \ldots$$
$$\int_{-\infty}^{-L_A}\int_{L}^{\infty}\int_{-L}^{L}\cdots\int_{-L}^{L} \mathcal{N}(\mathbf{x}_j; \mu_{\mathbf{x}_j}, \Sigma_{\mathbf{x}_j})\ d\mathbf{x}_j + \ldots$$
$$\underbrace{\int_{-\infty}^{-L_A}\int_{-\infty}^{-L}\int_{-L}^{L}\cdots\int_{-L}^{L}}_{j+2} \mathcal{N}(\mathbf{x}_j; \mu_{\mathbf{x}_j}, \Sigma_{\mathbf{x}_j})\ d\mathbf{x}_j$$

$$\mathbf{x}_j = \begin{bmatrix} x_k \\ \vdots \\ x_{k+j} \\ \hat{x}_{k+d|k} \end{bmatrix} \in \mathbb{R}^{j+2}$$

$\Sigma_{\mathbf{x}_j}(i_1, i_2) =$

$$\begin{cases} \mathbf{C}\mathbf{P}_{ss}^{L}\mathbf{C}^T + \mathbf{R} & 1 \leq i_1 = i_2 \leq j+1 \\ \mathbf{C}\mathbf{X}_{ss}\mathbf{C}^T & 1 \leq i_1 \neq i_2 \leq j+1 \\ \mathbf{C}\mathbf{A}^d(\mathbf{P}_{ss}^{L} - \hat{\mathbf{P}}_{ss}^{R})(\mathbf{A}^d)^T\mathbf{C}^T & i_1 = i_2 = j+2 \\ \mathbf{C}\mathbf{A}^d(\mathbf{P}_{ss}^{L} - \hat{\mathbf{P}}_{ss}^{R})(\mathbf{A}^{i_1 \vee i_2})^T\mathbf{C}^T & i_1 \neq i_2 : i_1 \vee i_2 = j+2 \end{cases}$$

$$\mathbf{X}_{ss} = \mathbf{A}^{i_1}\mathbf{P}_{ss}^{L}(\mathbf{A}^{i_2})^T + \mathbf{L}_{ss} - \mathbf{A}^{i_1}\mathbf{L}_{ss}(\mathbf{A}^{i_1})^T$$
$$\mathbf{L}_{ss} = \mathbf{A}\mathbf{L}_{ss}\mathbf{A}^T + \mathbf{A}^{i_2 - i_1}\mathbf{Q} \text{ where } i_1 < i_2$$

E. At least one up/downcrossing event within an interval

Finally, for this event, $P(C_{exact}, A_{approx}) = \sum_{j=1}^{d} P_{a_j}$, with identical definitions of P_{a_j}, $\mathbf{x}_j$, and $\Sigma_{\mathbf{x}_j}$ as in the previous case.

APPENDIX III
OPTIMAL ALARM SYSTEM COMPUTATIONS FOR $P(A_{approx})$ AND $P(C_{exact}, A_{approx})$

A. Up/downcrossing event spanning an interval

The probability of alarm, $P(A_{approx})$, for this event can be computed as in Eqn. III.1.

$$P(A_{approx}) = P(|\hat{x}_{k|k}| < L_A^-, |\hat{x}_{k+d|k}| > L_A^+) \qquad \text{(III.1)}$$
$$= \int_{-L_A^-}^{L_A^-}\int_{L_A^+}^{\infty} \mathcal{N}(\mathbf{x}; \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})\ d\mathbf{x} + \ldots$$
$$\int_{-L_A^-}^{L_A^-}\int_{-\infty}^{-L_A^+} \mathcal{N}(\mathbf{x}; \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})\ d\mathbf{x}$$

$$\mathbf{x} = \begin{bmatrix} \hat{x}_{k|k} \\ \hat{x}_{k+d|k} \end{bmatrix} \in \mathbb{R}^2$$

$$\Sigma_{\mathbf{x}} = \begin{bmatrix} \mathbf{C} \\ \mathbf{C}\mathbf{A}^d \end{bmatrix} (\mathbf{P}_{ss}^{L} - \hat{\mathbf{P}}_{ss}^{R}) \begin{bmatrix} \mathbf{C} \\ \mathbf{C}\mathbf{A}^d \end{bmatrix}^T$$

The probability, $P(C_{exact}, A_{approx})$, for this event is shown in Eqn. III.2.

$P(C_{exact}, A_{approx}) =$

$$P(|x_k| < L, |x_{k+d}| > L, |\hat{x}_{k|k}| < L_A^-, |\hat{x}_{k+d|k}| > L_A^+)$$

$$= \int_{-L}^{L}\int_{L}^{\infty}\int_{-L_A^-}^{L_A^-}\int_{L_A^+}^{\infty} \mathcal{N}(\mathbf{x}; \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})\ d\mathbf{x} + \ldots$$
$$\int_{-L}^{L}\int_{L}^{\infty}\int_{-L_A^-}^{L_A^-}\int_{-\infty}^{-L_A^+} \mathcal{N}(\mathbf{x}; \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})\ d\mathbf{x} + \ldots$$
$$\int_{-L}^{L}\int_{-\infty}^{-L}\int_{-L_A^-}^{L_A^-}\int_{L_A^+}^{\infty} \mathcal{N}(\mathbf{x}; \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})\ d\mathbf{x} + \ldots$$
$$\int_{-L}^{L}\int_{-\infty}^{-L}\int_{-L_A^-}^{L_A^-}\int_{-\infty}^{-L_A^+} \mathcal{N}(\mathbf{x}; \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})\ d\mathbf{x} \qquad \text{(III.2)}$$

$$\mathbf{x}_c \triangleq \begin{bmatrix} x_k \\ x_{k+d} \end{bmatrix}, \mu_{\mathbf{x}_c} = \begin{bmatrix} \mathbf{C}\mu_{\mathbf{q}} \\ \mathbf{C}\mu_{\mathbf{q}} \end{bmatrix} \in \mathbb{R}^2$$

$$\mathbf{x}_a \triangleq \begin{bmatrix} \hat{x}_{k|k} \\ \hat{x}_{k+d|k} \end{bmatrix}, \mu_{\mathbf{x}_a} = \begin{bmatrix} \mathbf{C}\mu_{\mathbf{q}} \\ \mathbf{C}\mu_{\mathbf{q}} \end{bmatrix} \in \mathbb{R}^2$$

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_c \\ \mathbf{x}_a \end{bmatrix}, \mu_{\mathbf{x}} = \begin{bmatrix} \mu_{\mathbf{x}_c} \\ \mu_{\mathbf{x}_a} \end{bmatrix} \in \mathbb{R}^4$$

$$\Sigma_{\mathbf{x}_c} = \begin{bmatrix} \mathbf{C}\mathbf{P}_{ss}^{L}\mathbf{C}^T + R & \mathbf{C}\mathbf{P}_{ss}^{L}(\mathbf{A}^T)^d\mathbf{C}^T \\ \mathbf{C}\mathbf{A}^d\mathbf{P}_{ss}^{L}\mathbf{C}^T & \mathbf{C}\mathbf{P}_{ss}^{L}\mathbf{C}^T + R \end{bmatrix}$$

$$\Sigma_{\mathbf{x}_a} = \begin{bmatrix} \mathbf{C} \\ \mathbf{C}\mathbf{A}^d \end{bmatrix} (\mathbf{P}_{ss}^{L} - \hat{\mathbf{P}}_{ss}^{R}) \begin{bmatrix} \mathbf{C} \\ \mathbf{C}\mathbf{A}^d \end{bmatrix}^T$$

$$\Sigma_{\mathbf{x}} = \begin{bmatrix} \Sigma_{\mathbf{x}_c} & \Sigma_{\mathbf{x}_a} \\ \Sigma_{\mathbf{x}_a} & \Sigma_{\mathbf{x}_a} \end{bmatrix}$$

B. End of interval up/downcrossing event

The probability of alarm, $P(A_{approx})$ and $P(C_{exact}, A_{approx})$ can be computed as in Eqns. III.3, and III.4, respectively.

$$\begin{aligned} P(A_{approx}) &= P(|\hat{x}_{k+d|k}| < L_A^-, |\hat{x}_{k+d+1|k}| > L_A^+) \\ &= \int_{-L_A^-}^{L_A^-}\int_{L_A^+}^{\infty} \mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})\ d\mathbf{x} + \ldots \\ &\quad \int_{-L_A^-}^{L_A^-}\int_{-\infty}^{-L_A^+} \mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})\ d\mathbf{x} \qquad \text{(III.3)} \\ \mathbf{x} &= \begin{bmatrix} \hat{x}_{k+d|k} \\ \hat{x}_{k+d+1|k} \end{bmatrix} \in \mathbb{R}^2 \\ \Sigma_{\mathbf{x}} &= \begin{bmatrix} \mathbf{C}\mathbf{A}^d \\ \mathbf{C}\mathbf{A}^{d+1} \end{bmatrix} (\mathbf{P}_{ss}^L - \hat{\mathbf{P}}_{ss}^R) \begin{bmatrix} \mathbf{C}\mathbf{A}^d \\ \mathbf{C}\mathbf{A}^{d+1} \end{bmatrix}^T \end{aligned}$$

and $P(C_{exact}, A_{approx}) = P(|x_{k+d}| < L, |x_{k+d+1}| > L, |\hat{x}_{k+d|k}| < L_A^-, |\hat{x}_{k+d+1|k}| > L_A^+)$

$$\begin{aligned} &= \int_{-L}^{L}\int_{L}^{\infty}\int_{-L_A^-}^{L_A^-}\int_{L_A^+}^{\infty} \mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})\ d\mathbf{x} + \ldots \\ &\quad \int_{-L}^{L}\int_{L}^{\infty}\int_{-L_A^-}^{L_A^-}\int_{-\infty}^{-L_A^+} \mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})\ d\mathbf{x} + \ldots \\ &\quad \int_{-L}^{L}\int_{-\infty}^{-L}\int_{-L_A^-}^{L_A^-}\int_{L_A^+}^{\infty} \mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})\ d\mathbf{x} + \ldots \\ &\quad \int_{-L}^{L}\int_{-\infty}^{-L}\int_{-L_A^-}^{L_A^-}\int_{-\infty}^{-L_A^+} \mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})\ d\mathbf{x} \qquad \text{(III.4)} \\ \mathbf{x}_c &\triangleq \begin{bmatrix} x_{k+d} \\ x_{k+d+1} \end{bmatrix}, \mu_{\mathbf{x}_c} = \begin{bmatrix} \mathbf{C}\mu_{\mathbf{q}} \\ \mathbf{C}\mu_{\mathbf{q}} \end{bmatrix} \in \mathbb{R}^2 \\ \mathbf{x}_a &\triangleq \begin{bmatrix} \hat{x}_{k+d|k} \\ \hat{x}_{k+d+1|k} \end{bmatrix}, \mu_{\mathbf{x}_a} = \begin{bmatrix} \mathbf{C}\mu_{\mathbf{q}} \\ \mathbf{C}\mu_{\mathbf{q}} \end{bmatrix} \in \mathbb{R}^2 \\ \mathbf{x} &= \begin{bmatrix} \mathbf{x}_c \\ \mathbf{x}_a \end{bmatrix}, \mu_{\mathbf{x}} = \begin{bmatrix} \mu_{\mathbf{x}_c} \\ \mu_{\mathbf{x}_a} \end{bmatrix} \in \mathbb{R}^4 \\ \Sigma_{\mathbf{x}_c} &= \begin{bmatrix} \mathbf{C}\mathbf{P}_{ss}^L\mathbf{C}^T + R & \mathbf{C}\mathbf{A}^d\mathbf{P}_{ss}^L(\mathbf{A}^T)^{d+1}\mathbf{C}^T \\ \mathbf{C}\mathbf{A}^{d+1}\mathbf{P}_{ss}^L\left(\mathbf{C}\mathbf{A}^d\right)^T & \mathbf{C}\mathbf{P}_{ss}^L\mathbf{C}^T + R \end{bmatrix} \\ \Sigma_{\mathbf{x}_a} &= \begin{bmatrix} \mathbf{C}\mathbf{A}^d \\ \mathbf{C}\mathbf{A}^{d+1} \end{bmatrix} (\mathbf{P}_{ss}^L - \hat{\mathbf{P}}_{ss}^R) \begin{bmatrix} \mathbf{C}\mathbf{A}^d \\ \mathbf{C}\mathbf{A}^{d+1} \end{bmatrix}^T \\ \Sigma_{\mathbf{x}} &= \begin{bmatrix} \Sigma_{\mathbf{x}_c} & \Sigma_{\mathbf{x}_a} \\ \Sigma_{\mathbf{x}_a} & \Sigma_{\mathbf{x}_a} \end{bmatrix} \end{aligned}$$

C. End of interval exceedance/fade event

The probability of alarm, $P(A_{approx})$ and $P(C_{exact}, A_{approx})$ can be computed as in Eqns. III.5, and III.6, respectively.

$$\begin{aligned} P(A_{approx}) &= P(|\hat{x}_{k+d|k}| \geq L_A^+) \qquad \text{(III.5)} \\ &= 2\Phi\left(-\frac{L_A^+}{\sqrt{\mathbf{C}\mathbf{A}^d(\mathbf{P}_{ss}^L - \hat{\mathbf{P}}_{ss}^R)(\mathbf{A}^d)^T\mathbf{C}^T}}\right) \end{aligned}$$

and

$$\begin{aligned} P(C_{exact}, A_{approx}) &= \\ &P(|x_{k+d}| > L, |\hat{x}_{k+d|k}| > L_A^+) \\ &\int_{L_A^+}^{\infty}\int_{L}^{\infty} \mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})\ d\mathbf{x} + \ldots \\ &\int_{-\infty}^{-L_A^+}\int_{L}^{\infty} \mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})\ d\mathbf{x} + \ldots \\ &\int_{L_A^+}^{\infty}\int_{-\infty}^{-L} \mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})\ d\mathbf{x} + \ldots \\ &\int_{-\infty}^{-L_A^+}\int_{-\infty}^{-L} \mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})\ d\mathbf{x} \qquad \text{(III.6)} \\ \mathbf{x} &= \begin{bmatrix} x_{k+d} \\ \hat{x}_{k+d|k} \end{bmatrix} \in \mathbb{R}^2 \\ \Sigma_{\mathbf{x}_j}(i_1,i_2) &= \begin{cases} \mathbf{C}\mathbf{P}_{ss}^L\mathbf{C}^T + \mathbf{R} & i_1 = i_2 = 1 \\ \mathbf{C}\mathbf{A}^d(\mathbf{P}_{ss}^L - \hat{\mathbf{P}}_{ss}^R)(\mathbf{A}^d)^T\mathbf{C}^T & \text{o.w.} \end{cases} \end{aligned}$$

D. At least one exceedance/fade event within an interval

The probability of alarm, $P(A_{approx})$ and $P(C_{exact}, A_{approx})$ can be computed as in Eqns. III.7, and III.8, respectively.

$$\begin{aligned} P(A_{approx}) &= P(\bigcup_{i=0}^{d} |\hat{x}_{k+i|k}| \geq L_{A_i}^+) \qquad \text{(III.7)} \\ &= 1 - P(\bigcap_{i=0}^{d} |\hat{x}_{k+i|k}| \leq L_{A_i}^+) \\ &= 1 - \underbrace{\int_{-L_{A_0}^+}^{L_{A_0}^+} \ldots \int_{-L_{A_d}^+}^{L_{A_d}^+}}_{d+1} \mathcal{N}(\mathbf{x},\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})d\mathbf{x} \\ \mathbf{x} &= \begin{bmatrix} \hat{x}_{k|k} \\ \vdots \\ \hat{x}_{k+d|k} \end{bmatrix}, \mu_{\mathbf{x}} = \begin{bmatrix} \mathbf{C}\mu_{\mathbf{q}} \\ \vdots \\ \mathbf{C}\mu_{\mathbf{q}} \end{bmatrix} \in \mathbb{R}^{d+1} \\ \Sigma_{\mathbf{x}} &= \begin{bmatrix} \mathbf{C} \\ \vdots \\ \mathbf{C}\mathbf{A}^d \end{bmatrix} (\mathbf{P}_{ss}^L - \hat{\mathbf{P}}_{ss}^R) \begin{bmatrix} \mathbf{C} \\ \vdots \\ \mathbf{C}\mathbf{A}^d \end{bmatrix}^T \end{aligned}$$

Mathematical curiosities of this type of covariance matrix as related to control theory, specifically the property of observability, are discussed in greater detail in [3]. Furthermore, it is convenient that the number of terms required to compute the complicated multi-dimensional level-crossing events as presented in Eqn. III.7 above can be reformulated to achieve better scaling properties. This is largely due to the way in which the aggregate probability computation can be rewritten using basic axioms of probability. Computing the same probability for the events related to thermal sensation complaint application shown in [3] requires the brute force inclusion/exclusion rule which results in an explosion of terms. Therefore, reformulation of the probability computation works advantageously, and this can also be applied to probability computations for events related to the thermal sensation complaint application. However, using the inclusion/exclusion rule

for these computations may provide a similar accuracy for less computational burden for reasons discussed in [3].

$$P(C_{exact}, A_{approx}) = P(C_{exact}) - \dots$$

$$P(|x_k| > L, \bigcap_{i=0}^{d} |\hat{x}_{k+i|k}| \leq L_{A_i}^+) - \dots$$

$$\sum_{j=1}^{d} P(\bigcap_{i=0}^{j-1} |x_{k+i}| < L, |x_{k+j}| > L, \bigcap_{i=0}^{d} |\hat{x}_{k+i|k}| \leq L_A^+) \tag{III.8}$$

where $P(C_{exact})$ was previously given in Eqn. 44, and $P(|x_k| > L, \bigcap_{i=0}^{d} |\hat{x}_{k+i|k}| \leq L_{A_i}^+) =$

$$\int_L^\infty \int_{-L_{A_0}^+}^{L_{A_0}^+} \dots \int_{-L_{A_d}^+}^{L_{A_d}^+} \mathcal{N}(\mathbf{x}, \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}}) d\mathbf{x} + \dots$$
$$\underbrace{\int_{-\infty}^{-L} \int_{-L_{A_0}^+}^{L_{A_0}^+} \dots \int_{-L_{A_d}^+}^{L_{A_d}^+} \mathcal{N}(\mathbf{x}, \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}}) d\mathbf{x}}_{d+2}$$

$$\mathbf{x} = \begin{bmatrix} x_k \\ \hat{x}_{k|k} \\ \vdots \\ \hat{x}_{k+d|k} \end{bmatrix}, \mu_{\mathbf{x}} = \begin{bmatrix} \mathbf{C}\mu_{\mathbf{q}} \\ \vdots \\ \mathbf{C}\mu_{\mathbf{q}} \end{bmatrix} \in \mathbb{R}^{d+2}$$

$$\Sigma_{\mathbf{x}_c} = \mathbf{C}(\mathbf{P}_{ss}^L - \hat{\mathbf{P}}_{ss}^R) \begin{bmatrix} \mathbf{C} \\ \vdots \\ \mathbf{C}\mathbf{A}^d \end{bmatrix}^T$$

$$\Sigma_{\mathbf{x}_a} = \begin{bmatrix} \mathbf{C} \\ \vdots \\ \mathbf{C}\mathbf{A}^d \end{bmatrix} (\mathbf{P}_{ss}^L - \hat{\mathbf{P}}_{ss}^R) \begin{bmatrix} \mathbf{C} \\ \vdots \\ \mathbf{C}\mathbf{A}^d \end{bmatrix}^T$$

$$\Sigma_{\mathbf{x}} = \begin{bmatrix} \mathbf{C}\mathbf{P}_{ss}^L\mathbf{C}^T + R & \Sigma_{\mathbf{x}_c} \\ \Sigma_{\mathbf{x}_c}^T & \Sigma_{\mathbf{x}_a} \end{bmatrix}$$

and

$$\sum_{j=1}^{d} \overbrace{P(\bigcap_{i=0}^{j-1} |x_{k+i}| < L, |x_{k+j}| > L, \bigcap_{i=0}^{d} |\hat{x}_{k+i|k}| \leq L_A^+)}^{P_{a_j}}$$

$$P_{a_j} = \Big(\int_{-L}^{L} \dots \int_{-L}^{L} \int_{L}^{\infty} \int_{-L_{A_0}^+}^{L_{A_0}^+} \dots \int_{-L_{A_d}^+}^{L_{A_d}^+} + \dots$$
$$\underbrace{\int_{-L}^{L} \dots \int_{-L}^{L} \int_{-\infty}^{-L} \int_{-L_{A_0}^+}^{L_{A_0}^+} \dots \int_{-L_{A_d}^+}^{L_{A_d}^+} \Big)}_{d+j+2} \mathcal{N}(\mathbf{x}, \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}}) d\mathbf{x}$$

$$\mathbf{x} = \begin{bmatrix} x_k \\ \vdots \\ x_{k+j} \\ \hat{x}_{k|k} \\ \vdots \\ \hat{x}_{k+d|k} \end{bmatrix}, \mu_{\mathbf{x}} = \begin{bmatrix} \mathbf{C}\mu_{\mathbf{q}} \\ \vdots \\ \mathbf{C}\mu_{\mathbf{q}} \end{bmatrix} \in \mathbb{R}^{d+j+2}$$

$$\Sigma_{\mathbf{x}_c}(i_1, i_2) = \begin{cases} \mathbf{C}\mathbf{P}_{ss}^L\mathbf{C}^T + \mathbf{R} & 1 \leq i_1 = i_2 \leq j+1 \\ \mathbf{C}\mathbf{X}_{ss}\mathbf{C}^T & 1 \leq i_1 \neq i_2 \leq j+1 \end{cases}$$

$$\mathbf{X}_{ss} = \mathbf{A}^{i_1}\mathbf{P}_{ss}^L(\mathbf{A}^{i_2})^T + \mathbf{L}_{ss} - \mathbf{A}^{i_1}\mathbf{L}_{ss}(\mathbf{A}^{i_1})^T$$

$$\mathbf{L}_{ss} = \mathbf{A}\mathbf{L}_{ss}\mathbf{A}^T + \mathbf{A}^{i_2-i_1}\mathbf{Q} \text{ where } i_1 < i_2$$

$$\Sigma_{\mathbf{x}_a} = \begin{bmatrix} \mathbf{C} \\ \vdots \\ \mathbf{C}\mathbf{A}^d \end{bmatrix} (\mathbf{P}_{ss}^L - \hat{\mathbf{P}}_{ss}^R) \begin{bmatrix} \mathbf{C} \\ \vdots \\ \mathbf{C}\mathbf{A}^d \end{bmatrix}^T$$

$$\Sigma_{\mathbf{x}_{ca}} = \begin{bmatrix} \mathbf{C} \\ \vdots \\ \mathbf{C}\mathbf{A}^j \end{bmatrix} (\mathbf{P}_{ss}^L - \hat{\mathbf{P}}_{ss}^R) \begin{bmatrix} \mathbf{C} \\ \vdots \\ \mathbf{C}\mathbf{A}^d \end{bmatrix}^T$$

$$\Sigma_{\mathbf{x}} = \begin{bmatrix} \Sigma_{\mathbf{x}_c} & \Sigma_{\mathbf{x}_{ca}} \\ \Sigma_{\mathbf{x}_{ca}}^T & \Sigma_{\mathbf{x}_a} \end{bmatrix}$$

E. At least one up/downcrossing event within an interval

The probability of alarm, $P(A_{approx})$ and $P(C_{exact}, A_{approx})$ can be computed as in Eqns. III.9, and III.10, respectively.

$$\begin{aligned} P(A_{approx}) &= P(|\hat{x}_{k|k}| \leq L_A^-, \bigcup_{i=1}^{d} |\hat{x}_{k+i|k}| \geq L_{A_i}^+) \quad \text{(III.9)} \\ &= P(|\hat{x}_{k|k}| \leq L_A^-) - \dots \\ &\quad P(|\hat{x}_{k|k}| \leq L_A^-, \bigcap_{i=1}^{d} |\hat{x}_{k+i|k}| \leq L_{A_i}^+) \\ &= \Phi\left(\frac{L_A^-}{\sqrt{\mathbf{C}(\mathbf{P}_{ss}^L - \hat{\mathbf{P}}_{ss}^R)\mathbf{C}^T}}\right) - \dots \\ &\quad \Phi\left(\frac{-L_A^-}{\sqrt{\mathbf{C}(\mathbf{P}_{ss}^L - \hat{\mathbf{P}}_{ss}^R)\mathbf{C}^T}}\right) - \dots \\ &\quad \underbrace{\int_{-L_A^-}^{L_A^+} \int_{-L_{A_1}^+}^{L_{A_1}^+} \dots \int_{-L_{A_d}^+}^{L_{A_d}^+} \mathcal{N}(\mathbf{x}, \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}}) d\mathbf{x}}_{d+1} \end{aligned}$$

$$\mathbf{x} = \begin{bmatrix} \hat{x}_{k|k} \\ \vdots \\ \hat{x}_{k+d|k} \end{bmatrix}, \mu_{\mathbf{x}} = \begin{bmatrix} \mathbf{C}\mu_{\mathbf{q}} \\ \vdots \\ \mathbf{C}\mu_{\mathbf{q}} \end{bmatrix} \in \mathbb{R}^{d+1}$$

$$\Sigma_{\mathbf{x}} = \begin{bmatrix} \mathbf{C} \\ \vdots \\ \mathbf{C}\mathbf{A}^d \end{bmatrix} (\mathbf{P}_{ss}^L - \hat{\mathbf{P}}_{ss}^R) \begin{bmatrix} \mathbf{C} \\ \vdots \\ \mathbf{C}\mathbf{A}^d \end{bmatrix}^T$$

And finally,

$$P(C_{exact}, A_{approx}) =$$

$$\sum_{j=1}^{d} P(\underbrace{\bigcap_{i=0}^{j-1} |x_{k+i}| < L, |x_{k+j}| > L}_{C_{exact}}, |\hat{x}_{k|k}| \leq L_A^-) - \dots$$

$$\sum_{j=1}^{d} P(C_{exact}(j), |\hat{x}_{k|k}| \leq L_A^-, \bigcap_{i=1}^{d} |\hat{x}_{k+i|k}| \leq L_A^+) \quad \text{(III.10)}$$

where $P(\bigcap_{i=0}^{j-1} |x_{k+i}| < L, |x_{k+j}| > L, |\hat{x}_{k|k}| \leq L_A^-) =$

$$\left(\int_{-L}^{L} \dots \int_{-L}^{L} \int_{L}^{\infty} \int_{-L_A^-}^{L_A^-} + \dots \right.$$

$$\left. \underbrace{\int_{-L}^{L} \dots \int_{-L}^{L} \int_{-\infty}^{-L} \int_{-L_A^-}^{L_A^-}}_{j+2} \right) \mathcal{N}(\mathbf{x}, \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}}) d\mathbf{x}$$

$$\begin{aligned}
\mathbf{x} &= \begin{bmatrix} x_k \\ \vdots \\ x_{k+j} \\ \hat{x}_{k|k} \end{bmatrix}, \mu_{\mathbf{x}} = \begin{bmatrix} \mathbf{C}\mu_{\mathbf{q}} \\ \vdots \\ \mathbf{C}\mu_{\mathbf{q}} \end{bmatrix} \in \mathbb{R}^{j+2} \\
\Sigma_{\mathbf{x}_c}(i_1, i_2) &= \begin{cases} \mathbf{C}\mathbf{P}_{ss}^L \mathbf{C}^T + \mathbf{R} & 1 \leq i_1 = i_2 \leq j+1 \\ \mathbf{C}\mathbf{X}_{ss}\mathbf{C}^T & 1 \leq i_1 \neq i_2 \leq j+1 \end{cases} \\
\mathbf{X}_{ss} &= \mathbf{A}^{i_1} \mathbf{P}_{ss}^L (\mathbf{A}^{i_2})^T + \mathbf{L}_{ss} - \mathbf{A}^{i_1} \mathbf{L}_{ss} (\mathbf{A}^{i_1})^T \\
\mathbf{L}_{ss} &= \mathbf{A}\mathbf{L}_{ss}\mathbf{A}^T + \mathbf{A}^{i_2 - i_1}\mathbf{Q} \text{ where } i_1 < i_2
\end{aligned}$$

$$\Sigma_{\mathbf{x}_a} = \mathbf{C}(\mathbf{P}_{ss}^L - \hat{\mathbf{P}}_{ss}^R) \begin{bmatrix} \mathbf{C} \\ \vdots \\ \mathbf{C}\mathbf{A}^d \end{bmatrix}^T$$

$$\Sigma_{\mathbf{x}} = \begin{bmatrix} \Sigma_{\mathbf{x}_c} & \Sigma_{\mathbf{x}_{ca}}^T \\ \Sigma_{\mathbf{x}_a} & \mathbf{C}(\mathbf{P}_{ss}^L - \hat{\mathbf{P}}_{ss}^R)\mathbf{C}^T \end{bmatrix}$$

and

$$P(C_{exact}, |\hat{x}_{k|k}| \leq L_A^-, \bigcap_{i=1}^{d} |\hat{x}_{k+i|k}| \leq L_A^+) =$$

$$\left(\int_{-L}^{L} \dots \int_{-L}^{L} \int_{L}^{\infty} \int_{-L_A^-}^{L_A^-} \int_{-L_{A_1}^+}^{L_{A_1}^+} \dots \int_{-L_{A_d}^+}^{L_{A_d}^+} + \dots \right.$$

$$\left. \underbrace{\int_{-L}^{L} \dots \int_{-L}^{L} \int_{-\infty}^{-L} \int_{-L_A^-}^{L_A^-} \int_{-L_{A_1}^+}^{L_{A_1}^+} \dots \int_{-L_{A_d}^+}^{L_{A_d}^+}}_{d+j+2} \right) \mathcal{N}(\mathbf{x}, \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}}) d\mathbf{x}$$

$$\begin{aligned}
\mathbf{x} &= \begin{bmatrix} x_k \\ \vdots \\ x_{k+j} \\ \hat{x}_{k|k} \\ \vdots \\ \hat{x}_{k+d|k} \end{bmatrix}, \mu_{\mathbf{x}} = \begin{bmatrix} \mathbf{C}\mu_{\mathbf{q}} \\ \vdots \\ \mathbf{C}\mu_{\mathbf{q}} \end{bmatrix} \in \mathbb{R}^{d+j+2} \\
\Sigma_{\mathbf{x}_c}(i_1, i_2) &= \begin{cases} \mathbf{C}\mathbf{P}_{ss}^L \mathbf{C}^T + \mathbf{R} & 1 \leq i_1 = i_2 \leq j+1 \\ \mathbf{C}\mathbf{X}_{ss}\mathbf{C}^T & 1 \leq i_1 \neq i_2 \leq j+1 \end{cases} \\
\mathbf{X}_{ss} &= \mathbf{A}^{i_1} \mathbf{P}_{ss}^L (\mathbf{A}^{i_2})^T + \mathbf{L}_{ss} - \mathbf{A}^{i_1} \mathbf{L}_{ss} (\mathbf{A}^{i_1})^T \\
\mathbf{L}_{ss} &= \mathbf{A}\mathbf{L}_{ss}\mathbf{A}^T + \mathbf{A}^{i_2 - i_1}\mathbf{Q} \text{ where } i_1 < i_2
\end{aligned}$$

$$\Sigma_{\mathbf{x}_a} = \begin{bmatrix} \mathbf{C} \\ \vdots \\ \mathbf{C}\mathbf{A}^d \end{bmatrix} (\mathbf{P}_{ss}^L - \hat{\mathbf{P}}_{ss}^R) \begin{bmatrix} \mathbf{C} \\ \vdots \\ \mathbf{C}\mathbf{A}^d \end{bmatrix}^T$$

$$\Sigma_{\mathbf{x}_{ca}} = \begin{bmatrix} \mathbf{C} \\ \vdots \\ \mathbf{C}\mathbf{A}^j \end{bmatrix} (\mathbf{P}_{ss}^L - \hat{\mathbf{P}}_{ss}^R) \begin{bmatrix} \mathbf{C} \\ \vdots \\ \mathbf{C}\mathbf{A}^d \end{bmatrix}^T$$

$$\Sigma_{\mathbf{x}} = \begin{bmatrix} \Sigma_{\mathbf{x}_c} & \Sigma_{\mathbf{x}_{ca}} \\ \Sigma_{\mathbf{x}_{ca}}^T & \Sigma_{\mathbf{x}_a} \end{bmatrix}$$

REFERENCES

[1] A. Svensson, "Event prediction and bootstrap in time series," Ph.D. dissertation, Lund Institute of Technology, September 1998.

[2] A. Svensson, J. Holst, R. Lindquist, and G. Lindgren, "Optimal prediction of catastrophes in autoregressive moving-average processes," *Journal of Time Series Analysis*, vol. 17, no. 5, pp. 511–531, 1996.

[3] R. A. Martin, "Optimal prediction, alarm, and control in buildings using thermal sensation complaints," Ph.D. dissertation, University of California, Berkeley, 2004.

[4] S.-I. Beckman, J. Holst, and G. Lindgren, "Alarm characteristics for a flood warning system with deterministic components," *Journal of Time Series Analysis*, vol. 11, no. 1, pp. 1–18, March 1987.

[5] J. DeMaré, "Optimal prediction of catastrophes with application to Gaussian processes," *Annals of Probability*, vol. 8, no. 4, pp. 840–850, August 1980.

[6] G. Lindgren, "Optimal prediction of level crossings in Gaussian processes and sequences," *Annals of Probability*, vol. 13, no. 3, pp. 804–824, August 1985.

[7] A. J. Rainal, "Zero-crossing principle for detecting narrow-band signals," *IEEE Transactions on Instrumentation and Measurement*, vol. IM-15, no. 1-2, pp. 38–43, 1966.

[8] M. Antunes, A. A. Turkman, and K. F. Turkman, "A Bayesian approach to event prediction," *Journal of Time Series Analysis*, vol. 24, no. 6, pp. 631–646, November 2003.

[9] R. Martin, "Unsupervised anomaly detection and diagnosis for liquid rocket engine propulsion," in *Proceedings of the IEEE Aerospace Conference*, Big Sky, MT, March 2007.

[10] T. H. Kerr, "False alarm and correct detection probabilities over a time interval for restricted classes of failure detection algorithms," *IEEE Transactions on Information Theory*, vol. IT-28, no. 4, pp. 619–631, July 1982.

[11] A. S. Willsky and H. L. Jones, "A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems," *IEEE Transactions on Automatic Control*, vol. 21, no. 1, pp. 108–112, 1976.

[12] F. Gustafson, "The marginalized likelihood ratio test for detecting abrupt changes," *IEEE Transactions on Automatic Control*, vol. 41, no. 1, pp. 66–78, 1996.

[13] T. H. K. III, "Drawbacks of residual-based event detectors like GLR or IMM filters in practical situations," *IEEE Transactions on Signal Processing (preprint)*, 2006.

[14] C. C. Federspiel, "Predicting the frequency of hot and cold complaints in buildings," *International Journal of HVAC&R Research*, vol. 6, no. 4, pp. 217–234, 2000.

[15] C. C. Federspiel, R. A. Martin, and H. Yan, "Thermal comfort models and "call-out" (complaint) frequencies," University of California, Berkeley, Center for the Built Environment, Tech. Rep., 2003.

[16] T. Fawcett, "ROC graphs: Notes and practical considerations for data mining researchers," Hewlett Packard Laboratories, Technical Report HPL-2003-4, 2003.

[17] A. Genz, "Numerical computation of multivariate normal probabilities," *Journal of Computational and Graphical Statistics*, vol. 1, pp. 141–149, 1992.

[18] T. H. K. III, "Integral evaluations enabling performance tradeoffs for two-confidence-region-based failure detection," *Journal of Guidance, Control, and Dynamics*, vol. 29, no. 3, pp. 757–762, May–June 2006.

[19] J. Larsen, L. K. Hansen, A. Szymkowiak, T. Christiansen, and T. Kolenda, "Webmining: Learning from the world wide web," *Computational Statistics and Data Analysis*, vol. 38, pp. 517–532, 2002. [Online]. Available: citeseer.ist.psu.edu/larsen01webmining.html

[20] N. H. Pontoppidan and J. Larsen, "Unsupervised condition change detection in large diesel engines," in *Proceedings of the 13th IEEE Workshop on Neural Networks for Signal Processing*. IEEE Press, September 2003, pp. 565–574.
[21] G. E. Forsythe, M. A. Malcolm, and C. B. Moler, *Computer Methods for Mathematical Computations*. Prentice-Hall, 1976.
[22] B. Kedem, *Time Series Analysis by Higher Order Crossings*. IEEE Press, 1994.
[23] S. O. Rice, "Mathematical analysis of random noise," *Bell System Technology Journal*, vol. 24, pp. 46–156, 1945.
[24] R. A. Martin, "Optimized response to thermal sensation complaints in buildings," Master's thesis, University of California, Berkeley, December 2000.
[25] C. C. Federspiel, R. A. Martin, and H. Yan, "Re-calibration of the complaint prediction model," *International Journal of HVAC&R Research*, vol. 10, no. 2, April 2004.
[26] Z. Ghahramani and G. E. Hinton, "Parameter estimation for linear dynamical systems," Department of Computer Science, University of Toronto, Technical Report CRG-TR-96-2, 1996.
[27] V. V. Digalakis, J. R. Rohlicek, and M. Ostendorf, "ML estimation of a stochastic linear system with the EM algorithm and its application to speech recognition," *IEEE Transactions on Speech and Audio Processing*, vol. 1, no. 4, pp. 431–442, 1993. [Online]. Available: citeseer.ist.psu.edu/digalakis92ml.html
[28] K. P. Murphy, "The Bayes' Net Toolbox for MATLAB," *Computing Science and Statistics*, vol. 33, 2001.
[29] S. D. Bay and M. Schwabacher, "Mining distance-based outliers in near linear time with randomization and a simple pruning rule," in *KDD '03: Proceedings of The Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY: ACM Press, 2003, pp. 29–38.
[30] D. L. Iverson, "Inductive system health monitoring," in *Proceedings of The 2004 International Conference on Artificial Intelligence (IC-AI04)*. Las Vegas, Nevada: CSREA Press, June 2004.
[31] H. Park, R. Mackey, M. James, M. Zak, M. Zynard, J. Sebghati, and W. Greene, "Analysis of Space Shuttle Main Engine Data Using Beacon-based Exception Analysis for Multi-Missions," in *Proceedings of the IEEE Aerospace Conference*, Big Sky, MT, March 2002.
[32] R. A. Martin, C. C. Federspiel, and D. M. Auslander, "Responding to thermal sensation complaints in buildings," *ASHRAE Transactions*, vol. 108, no. 1, pp. 407–412, Winter 2002.
[33] K. J. Åström, *Introduction to Stochastic Control Theory*. New York: Academic Press, 1970.
[34] K. J. Hunt, *Polynomial Methods in Optimal Control and Filtering*, ser. 49. Peter Peregrinus Ltd. on behalf of IEEE, 1993.
[35] G. F. Franklin, D. Powell, and M. L. Workman, *Digital Control of Dynamic Systems*, 2nd ed. Addison-Wesley Co., 1990.
[36] T. Norlander and P. M. Mäkilä, "A sampled-data theory for fixed structure LQ controller design," Luleå Technical University, Research Report 07, 2000.